GENOME-WIDE MAPPING OF FUNCTIONAL ELEMENTS IN GERM CELLS *IN VIVO*

NG JIA HUI (HUANG JIAHUI)

NATIONAL UNIVERSITY OF SINGAPORE

2012

GENOME-WIDE MAPPING OF FUNCTIONAL ELEMENTS IN GERM CELLS *IN VIVO*

NG JIA HUI (HUANG JIAHUI)

(B. Sc. (Hons.), NUS)

A THESIS SUBMITTED FOR THE DEGREE OF PHILOSOPHY

NUS GRADUATE SCHOOL FOR INTEGRATIVE SCIENCES AND ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2012

ACKNOWLEDGEMENTS

I would like to thank A*STAR for funding my PhD candidature, Dr Ng Huck Hui for his patient guidance and generous support, Dr Muratani Masafumi who established the small-scale ChIP protocol and taught me the technique, Dr Shyam Prabhakar and Dr Vibhor Kumar for bioinformatics analyses on the ChIP-Seq data, as well as Dr Thomas Lufkin and Dr Kraus Petra Marion for providing the E13.5 mouse embryos and help in animal work. In addition, I thank Ms Yaw Lai Ping for assisting me in the process of germ cell harvesting and IHC staining. I thank Dominic Heng Jien Chien for sharing Nr5a2-related reagents. I am also grateful to Dr Yu Hao, Dr Tara Huber and Dr Davor Solter, for their advice and support as members of my thesis advisory committee. I thank Kevin Andrew Uy Gonzales, Yeo Jia Chi and Dr Muratani Masafumi for their comments and suggestions on this thesis. I would also like to thank my colleagues from the Genome Institute of Singapore for sharing knowledge, imparting various technical skills and engaging in insightful discussions. Last but not the least, I thank my family members for their encouragements and moral support.

TABLE OF CONTENTS

1. Summary

The developmental progression of primordial germ cells (PGCs) to fully mature gametes involves extensive epigenetic changes which could be important for the gain of totipotency upon fertilization. Germ cells are also unique in that they can be reprogrammed to pluripotent embryonic germ cells (EGC) without the need for ectopic overexpression of any transcription factor if appropriate culture conditions are provided. Such plasticity to regain pluripotency or acquire totipotency could potentially be attributed to both transcriptional and epigenetic properties of germ cells, which are difficult to study on a genome-wide scale due to limited numbers of germ cells per embryo. In particular, it has been technically challenging to map histone modifications in germ cells as chromatin immunoprecipitation (ChIP) typically requires substantially high amounts of starting cell material. Using modified and optimized procedures, I purified germ cells from E13.5 male embryos, and performed chromatin immunoprecipitation (ChIP) coupled with a high-throughput sequencing technology. Genome-wide maps of H3K4me3, H3K27me3, H3K27ac and H2BK20ac have been generated for E13.5 male germ cells to enable the identification of active and poised regulatory elements, such as promoters and enhancers. In addition, expression profiling was performed on E11.5, E13.5 and E15.5 germ cells to study the transcriptional status/ changes of genes located near these regulatory elements, providing insight into the biological transitions that occur during this developmental period. Lastly, combining these datasets with motif enrichment analyses, potential germ cell regulators were identified and functionally validated *in vitro*. (238 words)

2. List of publications

Loh YH, Ng JH, Ng HH. Molecular framework underlying pluripotency. Cell Cycle. 2008 Apr 1;7(7):885-91.

Ng JH, Heng JC, Loh YH, Ng HH. Transcriptional and epigenetic regulation of embryonic stem cells. Mutat Res. 2008 Dec 1;647(1-2):52-8.

Feng B, Jiang J, Kraus P, Ng JH, Heng JC, Chan YS, Yaw LP, Zhang W, Loh YH, Han J, Vega VB, Cacheux-Rataboul V, Lim B, Lufkin T, Ng HH. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat Cell Biol. 2009 Feb;11(2):197-203.

Feng B, Ng JH, Heng JC, Ng HH. Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. Cell Stem Cell. 2009 Apr 3;4(4):301-12.

Heng JC, Feng B, Han J, Jiang J, Kraus P, Ng JH, Orlov YL, Huss M, Yang L,Lufkin T, Lim B, Ng HH. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. Cell Stem Cell. 2010 Feb 5;6(2):167-74.

Ng JH, Ng HH. LincRNAs join the pluripotency alliance. Nat Genet. 2010 Dec;42(12):1035-6.

Xue K, Ng JH, Ng HH. Mapping the networks for pluripotency. Philos Trans R Soc Lond B Biol Sci. 2011 Aug 12;366(1575):2238-46.

3. List of tables

4. List of figures

5. List of illustrations

6. List of symbols

Symbol	Definition
°C	degrees Celcius
µg	microgram
µl	microlitres
µM	micromolar
AP	alkaline phosphatase
bFGF	basic fibroblast growth factor
bp	base pairs
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
DMEM	Dulbecco's Modified Eagle's Medium
DNA	deoxyribonucleic acid
d.p.c.	days post coitum
EB	embryoid body
EGC	embryonic germ cell
ES cell (ESC)	embryonic stem cell
FACS	fluorescence-activated cell sorting
FBS	fetal bovine serum
GFP	green fluorescent protein
GO	gene ontology
h	hour
IHC	immunohistochemistry
iPSC	induced pluripotent stem cells
kb	kilobase
LIF	leukemia inhibitory factor
MEF	mouse embryonic fibroblast
mESC	mouse embryonic stem cell
mg	milligram
min	minutes
ml	millilitres
ng	nanogram
PCR	polymerase chain reaction
pg	picogram
qPCR	quantitative PCR
RA	retinoic acid
RNA	ribonucleic acid
rRNA	ribosomal RNA
s.e.m.	standard error of the mean
SCF	stem cell factor
WGA	whole genome amplification

7. Literature Review

7.1. Stem cells and developmental potency

Stem cells are characterized by an ability to self-renew and differentiate into a variety of cell types when provided with appropriate cues. Stem cells may be broadly categorized based on the developmental potency that they possess, from giving rise to only a single cell type (unipotency), multiple cell types of a particular lineage (multipotency), and all cell types of the embryo proper (pluripotency) (Jaenisch and Young, 2008). Totipotency refers to the ability to contribute to all lineages that are found in the embryo proper as well as extraembryonic tissues. In mammalian development, the zygote which forms upon fertilization as well as early blastomeres is totipotent. To date, there has been no report of totipotent stem cells that can be derived from vertebral organisms, cultured and maintained *in vitro*.

Under normal circumstances, embryonic development proceeds in a unidirectional manner from totipotent/ pluripotent cells that have broad developmental capacity to cells that have a specific lineage-restricted cell identity. However, recent studies have demonstrated that it is possible to coerce differentiated somatic cells to reacquire pluripotency *in vitro* (Takahashi and Yamanaka, 2006). This expansion of developmental potency or alteration of cell fate is generally termed as reprogramming. Reprogramming can be achieved by various techniques, namely cell fusion, somatic cell nuclear transfer and induced pluripotency (Yamanaka, 2007). More recently, numerous studies have reported that somatic cells can also be directly transdifferentiated from one somatic cell type to another (Jopling et al., 2011).

7.2. Sources of pluripotent stem cells

Early pluripotential studies originated from the observation that strain 129 mice have high incidence of testicular teratomas (Stevens and Little, 1954). It was later discovered that such teratocarcinoma properties could be attributed to a population of undifferentiated embryonal carcinoma (EC) cells (Kleinsmith and Pierce, 1964). Astonishingly, early mouse embryos when grafted into adult mice also resulted in teratocarcinoma formation (Solter et al., 1970; Stevens, 1970). Eventually in 1981, embryonic stem cells (ESCs) were directly derived from the inner cell mass (ICM) of mouse blastocysts (Evans and Kaufman, 1981; Martin, 1981). Unlike EC cells, mouse ESCs are karyotypically normal and give rise to germline chimeras (Bradley et al., 1984). This demonstration of germline transmission of ESCs has leapt the ESC field to one in which genetic manipulations and derivation of stable mouse transgenic lines are possible. Moreover, germline transmission implies that ESCs can contribute to functional gametes in addition to somatic cells of the three germ layers and thus are truly pluripotent by definition. Slightly more than a decade ago, human ESCs were derived (Thomson et al., 1998) and ignited much excitement about potential therapeutic applications of ESCs in regenerative medicine.

Besides ESC, other pluripotent cell types that include embryonic germ cells (EGCs) (Matsui et al., 1992; Resnick et al., 1992) and epiblast stem cells (EpiSCs) (Brons et al., 2007; Tesar et al., 2007) could be derived from early primordial germ cells (PGCs) and the post-implantation epiblast respectively. However, ESCs, EGCs and EpiSCs are all derived from early embryonic stages and thus not clinically feasible to obtain. The scarcity of human embryos, ethical concerns and advantages of autologous patient-specific pluripotent cells drives the motivation for newer methods

to derive pluripotent stem cells. Initial attempts through somatic cell nuclear transfer (Wilmut et al., 1997) and cell fusion (Cowan et al., 2005; Tada et al., 2001) were successful to a limited extent. Reprogramming via such methods was largely inefficient and generated new sets of problems such as developmental abnormalities (Tamashiro et al., 2002) and tetraploidy. In 2006, Yamanaka and colleagues reprogrammed mouse embryonic fibroblasts into induced pluripotent stem (iPS) cells via transduction with Oct4, Sox2, Klf4 and c-Myc (Takahashi and Yamanaka, 2006). Soon after, human iPS were derived using OCT4, SOX2, KLF4 and c-MYC (Park et al., 2008; Takahashi et al., 2007) or with the latter two factors replaced by NANOG and LIN28 (Yu et al., 2007).

Following the pilot discovery of iPS, numerous groups have quickly replicated and improved the reprogramming technique. Thus far, iPS cell lines have been generated from various cell types including hepatocytes (Aoi et al., 2008), gastric epithelial cells (Aoi et al., 2008), neural progenitor cells (Mikkelsen et al., 2008), pancreatic β cells (Stadtfeld et al., 2008), lymphocytes (Hanna et al., 2008) and keratinocytes (Aasen et al., 2008). Efforts were also made to avoid genomic integration of the transgenes by introducing reprogramming factors as purified proteins (Kim et al., 2009a; Zhou et al., 2009) or RNA (Warren et al., 2010) instead of through viral transduction. Notably, the complexity of reprogramming factor cocktail and efficiency of reprogramming depends on the intrinsic characteristics of the starting cell type. For example, while fibroblasts generally require three reprogramming factors if c-Myc is excluded (Nakagawa et al., 2008), neural stem cells can be reprogrammed to iPSC with Oct4 alone (Kim et al., 2009b; Nakagawa et al., 2008). In addition, reprogramming efficiencies could be improved using chemical inhibitors of DNA methylases, histone

deacetylases and histone methylases (Feng et al., 2009b), suggesting that epigenetic alteration is also an important element of the reprogramming process.

7.3. Mouse embryonic stem cells

7.3.1. The core pluripotency factors and extended transcriptional network

The ability to maintain ESC pluripotency and to direct de-differentiation during iPS reprogramming stems from our understanding of the molecular circuitry that underlies pluripotency. In ESC, the core transcriptional network comprises Oct4 (encoded by the *Pou5f1* gene), Sox2 and Nanog which result in pronounced phenotypes upon knockout *in vivo*. Disruption of *Pou5f1* caused the ICM to aberrantly differentiate into the trophectoderm lineage instead of the embryo proper (Nichols et al., 1998); *Nanog*-null ICM failed to develop into the epiblast (Silva et al., 2009); *Sox2* knockout mouse resulted in an early embryonic lethal phenotype (Avilion et al., 2003). Together, Oct4, Sox2 and Nanog promote pluripotency and self-renewal while suppressing differentiation (Boyer et al., 2005; Loh et al., 2006). In recent years, the ESC transcription network rapidly expanded to include other transcription factors such as Klfs (Jiang et al., 2008), Sall4 (Wu et al., 2006b), Zic3 (Lim et al., 2007; Wu et al., 2006b), Tcf3 (Cole et al., 2008), Znf143 (Chen et al., 2008a) and Ronin (Dejosez et al., 2008). Large-scale or later genome-wide RNA-interference (RNAi) screens also led to new additions to the pluripotency network such as Esrrb, Tcf3, Tcl1 (Ivanova et al., 2006), Paf1c (Ding et al., 2009), Cnot3, Trim28a (Hu et al., 2009) and the mediator/ cohesin complexes (Kagey et al., 2010).

To gain insight into how these pluripotency factors maintain ESCs in a continually self-renewing yet undifferentiated state, it is helpful to identify their downstream target genes. Chromatin immunoprecipitation (ChIP)-based technologies have enabled the genome-wide mapping of binding sites of core pluripotency factors Oct4, Sox2 and Nanog as well as other pluripotency-associated transcription factors and co-regulators in mouse ESCs (mESCs) (Chen et al., 2008b; Kim et al., 2008). Co-localization analyses revealed that pluripotency transcription factors can be clustered into two major groups: an Oct4-centric module (Oct4, Sox2, Nanog, Smad1, STAT3, Dax1, Nac1, Zfp281, Esrrb, Nr5a2, Tcfcp2l1 and Klf4) and a Myc-centric module (c-Myc, n-Myc, E2F1, Zfx, Rex1 and Ronin). The Oct4-centric cluster co-localizes extensively with p300 which is commonly enriched at enhancer regions, and directly links the transcription network with the LIF/STAT3 and BMP4/Smad1 signaling pathways (Chen et al., 2008b). It was also shown that the strength of transcription regulation positively correlates with the number of transcription binding at the same site (Kim et al., 2008). Downstream target genes include a broad spectrum of molecules such as transcription factors, microRNAs, lincRNAs, chromatin-remodelling and histone modifying proteins (Ng and Surani, 2011). Besides co-regulating common target genes, the expanded repertoire of pluripotency transcription factors also autoregulate themselves and each other (Ng and Surani, 2011), contributing to a stabilized yet sensitive transcriptional network in ESCs.

7.3.2. Epigenetic features of mouse ESCs

Multiple lines of evidence suggest that the pluripotency transcriptional network is intricately intertwined to epigenetic regulation. For example, Oct4, Sox2 and Nanog

co-bind and co-regulate epigenetic factors such as *Smarcad1*, *Myst3*, *Eset*, *Eed*, *Suz12*, and *Phc1* (Endoh et al., 2008; Loh et al., 2006). The interaction between transcription factors and epigenetic regulators is bi-directional such as the case whereby Oct4 binds to the *Jmjd1a* and *Jmjd2c* genes, which encode for H3K9 demethylases that in turn regulate the expression of pluripotent genes such as *Nanog* and *Tcl1* in mESCs (Loh et al., 2007). Besides H3K9 methylation, DNA methylation is also an important epigenetic mark to silence pluripotency genes such as *Pou5f1*, *Nanog*, *Lefty1* and *Tdgf1* (Farthing et al., 2008), and must be actively erased during reprogramming by an activation-induced cytidine deaminase (AID)-mediated mechanism (Bhutani et al., 2010). On this note, it is rather surprising that global CpG methylation patterns are relatively similar between mESC and differentiated cell types (Meissner et al., 2008).

In general, the chromatin state of mESCs can be described as hyperdynamic and structurally relaxed (Meshorer and Misteli, 2006). Mouse ESCs also have higher levels of H3K9ac than somatic cells, corresponding to an open chromatin that is transcriptionally permissive (Mattout and Meshorer, 2010). In addition, active H3K4me3 and repressive H3K27me3 epigenetic marks were found to co-localize at highly conserved noncoding elements (HCNEs) in mouse ESCs (Azuara et al., 2006; Bernstein et al., 2006). The term bivalent is now used to describe such H3K4me3 and H3K27me3-marked regions, which are often enriched in genes with developmental functions and are expressed at low levels in ESCs. Initially thought as an ESC-specific feature, bivalent domains were later shown to exist in other differentiated cell types as well (Mikkelsen et al., 2007). It is postulated that bivalent marks keep

developmental genes poised for activation and they resolve to either H3K4me3 or H3K27me3 upon differentiation.

7.4. Germ cells: specification, migration and differentiation

At the proximal posterior epiblast, PGC precursors receive bone morphogenetic protein (BMP) signals from the surrounding extraembryonic tissues. At E6.25, a few cells begin expressing Blimp1 which represses *Hoxb1* and other mesodermal genes (Ohinata et al., 2005). These *Blimp1*-positive cells represent lineage-restricted PGCs that eventually give rise to about 40 *Stella*-positive PGCs (Ohinata et al., 2005) at E7.25 (Illustration 1). Soon after at around E7.5, PGCs begin migration from the proximal epiblast to eventually arrive at the developing gonad at around E10.5 to E11.5 (Richardson and Lehmann, 2010). During migration, PGCs continuously proliferate and undergo global epigenetic changes (Sasaki and Matsui, 2008; Tam and Snow, 1981). Cells that fail to arrive at the gonad would undergo programmed cell death to safeguard against development of germline teratomas and this process involves factors such as Dnd, p53 and Bax (Richardson and Lehmann, 2010). Over the next two days, the post-migratory germ cells receive signals from the neighboring gonadal somatic cells (Sertoli and Leydig in male gonads; Granulosa cells in female gonads), undergo sex determination and commit to either oogenesis or spermatogenesis (Kocer et al., 2009). By E12.5-E13.5, testis cords can be visualised in the male gonads. At E13.5, male germ cells undergo mitotic arrest while female germ cells initiate meiosis. These sex-dependent cell cycle changes depend on both signaling from gonadal somatic cells as well as activation of intrinsic factors that

suppress meiosis and/or initiate differentiation of male germ cells, such as Nanos2, Dazl and Ddx4 (Lin and Page, 2005; Suzuki and Saga, 2008; Tanaka et al., 2000).

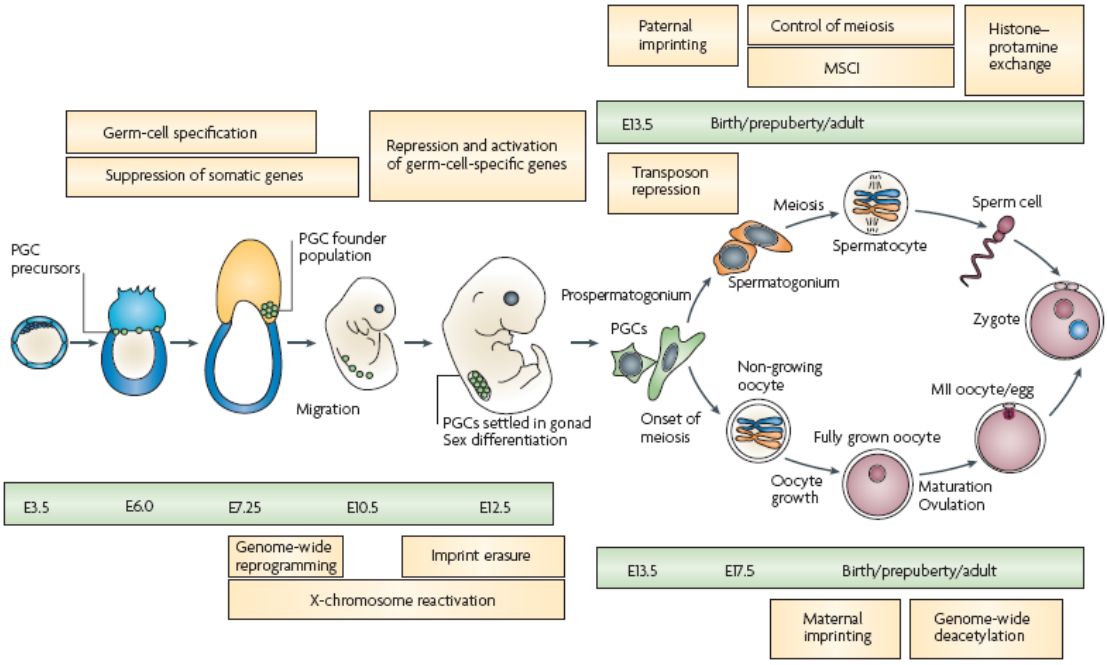

Illustration 1: Germ cell development. Chronology of major events in germ cell development that include specification, migration, sex determination and sex-specific differentiation. Adapted from Sasaki and Matsui, Nature Review Genetics 2008.

7.4.1. Transcriptional network

Besides Blimp1, which specifies germ cell commitment by suppressing differentiation to other lineages (Ohinata et al., 2005), *Prdm14* is also expressed in early PGCs at E6.75 (Yamaji et al., 2008). *Prdm14*-null mice are sterile due to a defect in germ cell specification. However, unlike Blimp1, Prdm14 does not repress the somatic

mesodermal program but instead is essential for activating pluripotency-associated factors and epigenetic reprogramming in germ cells (Yamaji et al., 2008).

Following gastrulation from about 8.5 d.p.c onward, *Pou5f1* (encoding the Oct4 protein) expression is confined to germ cells (Scholer et al., 1990). In males, *Pou5f1* expression is maintained until the formation of type A spermatogonia after birth. In females, *Pou5f1* is downregulated when female germ cells enter meiosis and reactivated upon oocyte maturation (Pesce et al., 1998). The germ cell-specific expression of *Pou5f1* is striking given the well-established roles of Oct4 as a pluripotency factor in the early embryos and mESCs (Nichols et al., 1998; Niwa et al., 2000), as well as its role in inducing reprogramming of somatic cells (Takahashi and Yamanaka, 2006). Notably, germ cell-specific depletion of *Pou5f1* resulted in an apoptotic response between 9.5 to 10.5 d.p.c, suggesting that Oct4 may be involved in regulating PGC survival (Kehler et al., 2004). The cell death phenotype, rather than differentiation, also indicates that Oct4's functions in germ cells may differ from that in the ICM (Kehler et al., 2004).

Other than *Pou5f1*, another core pluripotency factor *Nanog* is also expressed in PGCs. Initially repressed in E7.25-E7.5 PGCs, *Nanog* expression becomes detectable in migratory PGCs of both sexes and subsequently repressed again in E13.5 female germ cells that begin to undergo meiosis and in E14.5-E16.5 male germ cells that are undergoing mitotic arrest (Yamaguchi et al., 2005). *Nanog*-null PGCs failed to progress beyond E11.5, and hence Nanog is functionally implicated in the execution of germ cell development after PGCs arrive at the gonad (Chambers et al., 2007).

Knowledge of the transcriptional network of PGCs and post-migratory germ cells is rather limited to a handful of factors due to several reasons: the inaccessibility and limited numbers of germ cell per embryo, heterogeneity and dynamic nature of germ cells from pre-migratory to post-migratory, as well as the lack of a robust and convenient *in vitro* model system. While *in vitro* studies may not provide an exact reflection of *in vivo* events, they provide useful platforms for screening novel regulators of germ cells. For example, West et al. performed RNAi-mediated knockdown of candidate genes on *in vitro* generated germ cells. Due to a lack of a marker that is specific to early germ cells but not expressed in ESCs, potential germ cell regulators have to be indirectly screened by quantifying the number of EGC colonies formed after culture-mediated conversion (West et al., 2009). Using this method, they discovered a role of Lin28 in germ cell formation *in vitro* and further validated that *Lin28* knockdown reduced germ cell numbers by E12.5.

7.4.2. Epigenetic reprogramming in germ cells

Germ cells undergo extensive epigenetic reprogramming during and after their migration into the genital ridges (Illustration 2). These epigenetic changes are pivotal for parental imprint erasure and X-chromosome reactivation, and potentially prime germ cells for the reacquisition of totipotency (Sasaki and Matsui, 2008). In migratory germ cells, H3K9 methyltransferase GLP (also known as Ehmt1) is downregulated at E7.25 followed by a reduction of the repressive H3K9me2 mark which reaches very low levels by E8.75 (Seki et al., 2005; Seki et al., 2007). At around the same time, H3K27me3 signals are strongly elevated at E8.5-E9.0 and remain at relatively high levels as compared to neighboring somatic cells at least until E12.5 (Seki et al., 2005).

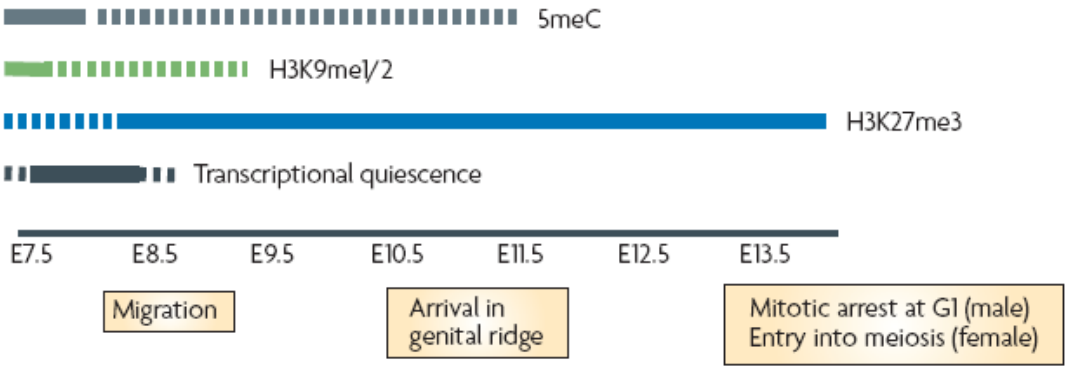

Illustration 2: Epigenetic reprogramming in PGCs. Dashed bars indicate lower levels of the specified epigenetic modifications as compared to solid bars. Adapted from Sasaki and Matsui, Nature Review Genetics 2008.

During the period when H3K9me2 is present at a low level and when H3K27me3 is not yet elevated, PGCs exhibit repression of RNA polymerase II-dependent transcription (Seki et al., 2007). Other than changes in histone modifications, germ cells also undergo two rounds of DNA demethylation at E8.0 and when germ cells arrive at the genital ridges at around E12.5 (Seki et al., 2005). By E13.5, both male and female germ cells are exceedingly hypomethylated with DNA methylation reaching levels that are even lower than methylation-deficient mESCs (Popp et al., 2010). Global DNA demethylation in germ cells between E8.0 to E13.5 have been associated with passive mechanisms such as absence of the maintenance methyltransferase Dnmt1 at around E8.0 accompanied by low levels of *de novo* methyltransferase Dnmt3a and Dnmt3b, as well as active mechanisms that are dependent on the cytidine deaminase AID (Popp et al., 2010; Seki et al., 2005). Overall, prior to E13.5, germ cells reduce stable repressive marks such as H3K9me2

and DNA methylation, and gain H3K27me3 which is a repressive mark characteristic of bivalent domains and is associated with greater plasticity. It is tantalizing to speculate that such "permissive" epigenetic landscape of germ cells may explain why early germ cells readily convert to pluripotent EGCs in culture. After E13.5, epigenetic events include the re-establishment of imprints after mitotic arrest for male germ cells and during post-natal oocyte growth for female germ cells (Reik et al., 2001). H3K4 methylation by Prdm9 was also shown to be necessary for meiosis in both male and female germ cells (Hayashi et al., 2005).

7.4.3. Conversion of germ cells to pluripotent stem cells

Primordial germ cells are unipotent cells which are committed to the germ lineage, eventually differentiating into male or female gametes. The fusion of the mature male and female gametes, also known as the sperm and oocyte, will generate a totipotent zygote which has the capacity to give rise to all cell types of the embryo proper as well as extra-embryonic cell types. Apart from its natural ability to regain totipotency upon fertilization, PGCs may be reprogrammed by *in vitro* culture conditions to generate pluripotent embryonic germ cell lines (Matsui et al., 1992; Resnick et al., 1992). Unlike the unipotent PGCs, pluripotent EGCs can give rise to germline chimeras when injected into the blastocyst (Labosky et al., 1994; Stewart et al., 1994).

To date, germ cells are the only cell type that can be reverted back to a pluripotent stem cell simply by providing the appropriate culture system. Remarkably, this conversion does not require the overexpression of any transcription factor and occurs relatively rapidly in 7 to 10 days (Durcova-Hills et al., 2008; Matsui et al., 1992; Resnick et al., 1992). Early experimentations led by two independent groups converge

on the same cocktail of three culture supplements that are essential for the reversion of unipotent PGCs to pluripotent EGCs: leukemia inhibitory factor (LIF), stem cell factor (SCF) and basic fibroblast growth factor (bFGF) (Matsui et al., 1992; Resnick et al., 1992). Both LIF and SCF promote survival and proliferation of germ cells (Dolci et al., 1991; Godin et al., 1991; Matsui et al., 1991; Pesce et al., 1993), but together are insufficient to sustain germ cell culture beyond 7 days (Donovan et al., 1986; Matsui et al., 1991). The inclusion of bFGF is critical for the formation of tightly packed alkaline phosphatase (AP)-positive EGC colonies that resemble mESC (Matsui et al., 1992; Resnick et al., 1992). It was later shown that the effect of bFGF can also be achieved with Trichostatin A (TSA) which is an inhibitor of histone deacetylases (Durcova-Hills et al., 2008). Both bFGF and TSA result in the downregulation of *Blimp1* and hence de-repression of *c-Myc* and *Klf4* (Durcova-Hills et al., 2008). Peculiarly, the efficiency of EGC conversion drastically declines as germ cells progress from E8.5 to E12.5 (Labosky et al., 1994; Matsui et al., 1992). For EGCs that were obtained from E12.5 embryos, all were of male origin and none were from female gonads (Labosky et al., 1994). Therefore, it is likely that early migratory PGCs differ from post-migratory gonadal germ cells in terms of responsiveness to growth factors in culture, and this change might be more prominent for female germ cells by E12.5.

7.5. Histone modifications and regulatory elements

Comparative analyses of genome sequencing data revealed that only less than half of the evolutionary conserved regions lie within protein-encoding regions (Chinwalla, 2002; Siepel et al., 2005). Among these highly conserved non-coding regions lie

functional regulatory elements such as promoters and enhancers. Located immediately upstream of transcriptional start sites (TSS), promoters are sites bound and regulated by both transcription factors and the transcriptional initiation machinery (Visel et al., 2009b). Enhancers are characteristically bound of a high density of transcription factors, and activate promoters by the recruitment of coactivators which leads to DNA looping (Visel et al., 2009b). However, unlike promoters, enhancers can be located distant from the TSS and function in an orientation-independent manner.

Over the years, numerous studies have shown that the combinatorial occurrence of specific histone modifications is predictive of the type and activity of regulatory elements. For example, H3K4me3 was found to be enriched at promoters, but it's correlation to expression of downstream genes depends on the GC content of the promoter region (Mikkelsen et al., 2007). Enhancers are characterized by lower levels of H3K4me3, higher levels of H3K4me1 and binding by p300 (Heintzman et al., 2007; Visel et al., 2009a). As compared to promoters, activation of enhancers exhibited greater cell-type specificity (Heintzman et al., 2009; Visel et al., 2009a). Among the histone modifications, H2BK20ac (Shyam et al. unpublished data) and H3K27ac (Creyghton et al., 2010; Rada-Iglesias et al., 2011) have been associated with active promoters/ enhancers while H3K27me3 occurs at bivalent promoters/ enhancers (Azuara et al., 2006; Bernstein et al., 2006) that are poised for activation or repression during development/ differentiation.

7.6. ChIP-Seq and its technical limitations

ChIP is a powerful method for studying the association of transcription factors or chromatin-associated factors to DNA. The presence of histone modifications at nucleosomes can also be detected with ChIP. Succinctly, ChIP involves the enrichment of fragmented crosslinked protein-DNA complexes by using an antibody that specifically pulls down the protein/ histone mark of interest. Following de-crosslinking, the purified DNA fragments can then be quantified with qPCR (ChIP-qPCR), microarray (ChIP-Chip) or high-throughput sequencing (ChIP-Seq) (Illustration 3). ChIP-Seq surpasses ChIP-Chip in several aspects such as higher resolution, flexibility in sequencing depth and greater genomic coverage (Park, 2009). However, success in ChIP-Seq is still dependent on antibody quality and sample quantity. Sequencing costs, sequencing artifacts/ bias and massive data management are also issues of concern for ChIP-Seq.

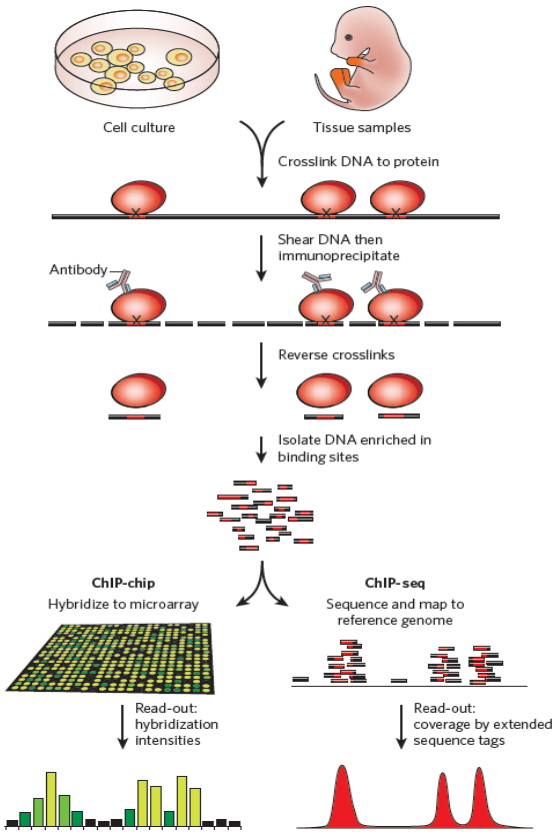

Illustration 3: Schematic overview of ChIP-Chip and ChIP-Seq.

The crosslinked DNA-protein complexes from either cell culture or tissue samples are sheared followed by immunoprecipitation with an antibody that binds the protein of interest. After reversal of crosslinks, enrichment of DNA fragments can be detected by hybridization to microarrays (ChIP-Chip) or coupled to massively parallel sequencing (ChIP-Seq). Adapted from Visel et al., Nature 2009.

Although the amount of DNA input required for ChIP-Seq has been significantly reduced as compared to the antecedent ChIP-Chip method, at least 10-50 ng of immunoprecipitated DNA is still needed for each sample (Park, 2009) and this would translate to a typical starting material of millions of cells. Unlike cell culture systems that provide abundant material for ChIP, it is challenging to achieve similar amounts of ChIP-DNA from small cell populations such as cell types that can only be purified from *in vivo*.

7.7. Current knowledge gap and research objectives

Given the unique "plasticity" and extensive epigenetic reprogramming of germ cells, it would be informative to map histone modifications on a genome-wide manner. Due to technical challenges such as limited numbers of germ cells, previous studies have largely only analyzed epigenetic changes on a global level by immunostaining and western blots (Seki et al., 2005; Seki et al., 2007), and the chromatin state at precise genomic locations has remained elusive.

As it is neither feasible nor cost-effective to purify germ cells on such as large-scale, it is necessary to modify and optimize the germ cell isolation and ChIP procedures for small sample quantities. In this project, I will demonstrate the feasibility of a modified small-scale ChIP-Seq procedure and use the optimized method to map several histone modifications including H3K4me3, H3K27me3, H3K27ac and H2BK20ac in purified fetal male germ cells. These histone modification data would elucidate the epigenetic landscape in germ cells as well as reveal genomic locations and activity status of regulatory elements (promoters and enhancers) that are active, repressed and poised in germ cells. Going further, motif enrichment analyses would also be performed on the

germ cell regulatory elements to uncover potential regulators of germ cell development. In addition to ChIP-Seq, I will perform microarray to profile transcriptional changes in fetal germ cells to gain insight into germ cell development.

8. Material and Methods

8.1. Purification of germ cells

Gonads were microscopically dissected from E13.5 embryos, which were derived from the intercross between male *Pou5f1*-GFP male mice (Jackson's lab, stock no. 004654) and female wild type CD1 mice. Gonads were treated with 0.25% Trypsin-EDTA (Gibco) for 10-15 min at 37°C, with occasional pipetting to aid dissociation, and trypsin was inactivated by addition of 15% Fetal Bovine Serum (FBS)/ DMEM (Gibco). Cells were filtered through 40 μm cell strainers (BD Falcon) to obtain single cell suspension for FACS sorting (FACSAria II SORP, BD Bioscience).

8.2. ChIP assay

ChIP assays were performed as previously described (Loh et al., 2006). Cells were crosslinked with 1% formaldehyde for 8 min at room temperature and the formaldehyde was quenched with 125 mM glycine. FACS-purified cells were fixed, quenched and rinsed with TBSE buffer twice before storage at -80°C until sufficient cell numbers are accumulated for ChIP experiments. Chromatin was sonicated using the bioruptor (Diagenode) and chromatin extracts were immunoprecipitated with anti-H3K4me3 (04-745, Millipore), anti-H3K27me3 (07-449, Millipore), anti-H3K27ac (ab4729, Abcam) and H2BK20ac (ab52988, Abcam) antibodies. Following de-crosslinking, small-scale ChIP samples were subjected to an additional 15 cycles of amplification using the GenomePlex Single Cell Whole Genome Amplification Kit (WGA4, Sigma), and further purified using the QIAquick PCR purification kit. For all ChIP experiments, quantitative PCR (qPCR) analyses were performed in real-time

using the ABI PRISM 7900 Sequence Detection System and SYBR Green Master Mix (ABI). Relative occupancy values were determined by calculating the immunoprecipitation efficiency (ratios of the amount of immunoprecipitated DNA to that of the input sample). Small-scale ChIP samples were subjected to two rounds of BpmI restriction digestion to excise PCR adaptors prior to ChIP-Seq library preparation. For ChIP-Seq library preparation (ChIP-Seq DNA Sample Prep Kit IP-102-1001, Illumina), 5-15 ng of ChIP DNA (as quantified by Picogreen assay) was used for end-repair and adaptor ligation followed by 15 cycles of amplification. 200-300 bp size fragments were selectively cut from the agarose gel and purified using the Qiagen gel extraction kit. The extracted DNA was subjected to ChIP-Seq sequencing for 36 cycles according to the manufacturer's instructions (Illumina).

8.3. Microarray sample preparation from small quantities of RNA

RNA was extracted from FACS-purified cells using the Arcturus PicoPure RNA Isolation Kit (Applied Biosystems) with on-column DNase treatment to eliminate potential DNA contamination. RNA quality was assessed using the RNA 6000 Pico LabChip Kit (Agilent) and quantitated using the Quant-iT RiboGreen RNA Reagent (Invitrogen). 10 ng of good-quality RNA, as defined by RIN value greater than 9.0, is amplified using the Ovation RNA Amplification System V2 (NuGEN) and purified using the Zymo Research DNA Clean & Concentrator-25 kit. Biotin-labelling was performed by incubating 5 ug of the amplified cDNA with UNG enzyme (Epicentre Biotechnologies) at 50°C for 30 min. Labelling buffer (0.952 M acetic acid and 28 mM $MgCl_2$) and ARP solution (Molecular Probes) was added and the mixture was

incubated at 50°C for 1 h, followed by a final purification using the Zymo Research DNA Clean & Concentrator-25 kit.

1.5 μg of biotin-labelled cDNA was hybridized to microarrays (MouseWG-6 Expression BeadChip version 2.0) at 48°C, according to NuGEN's recommendation. Subsequent microarray chip processing was performed according to Illumina's instructions and the arrays were scanned with the Illumina microarray platform. Three biological replicate microarray data was generated for E11.5, E13.5 and E15.5 germ cells and E13.5 GFP-negative cells were used as control. Differentially expressed genes were selected based on Significance Analysis of Microarrays (SAM) (Stanford University) using the following criteria: detection probability greater than 0.95; fold change (FC) < 0.6 for downregulated genes, FC > 1.5 for upregulated genes and q-value < 1%. Expression heatmaps were plotted using the Cluster 3.0 software and Java TreeView (Stanford University) and gene ontology analysis were performed using DAVID Bioinformatics Resources 6.7 (NIH) (Huang et al., 2008, 2009).

8.4. Cell culture

E14 mESCs were cultured on gelatin-coated dishes in Dulbecco's modified Eagle medium (DMEM, GIBCO), supplemented with 15% heat-inactivated fetal bovine serum (FBS,GIBCO), 0.055 mM β-mercaptoethanol (GIBCO), 2 mM L-glutamine (GIBCO), 0.1 mM MEM nonessential amino acid (GIBCO), 20 μg ml^{-1} gentamicin (GIBCO) and 1,000 units ml^{-1} of LIF (Chemicon). Mouse ESCs were routinely passaged every 2 - 3 days at a splitting ratio of 1:8. v6.4 mouse ESCs and established stable cell lines were cultured on mitomycin C-treated MEF feeders in the same ES cell medium and passaged every 2 - 3 days. MEFs were isolated from E13.5 embryos

by dissociation in 0.05% trypsin at 37 ºC for 10 min and cultured in 15% FBS-DMEM containing 20 μg ml^{-1} gentamicin. 293T cells were cultured in 10% FBS-DMEM containing 20 μg ml^{-1} gentamicin.

8.5. Transient transfection of mouse ESCs

Transfection of E14 mESCs with knockdown or overexpression plasmids was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Briefly, 2 μg of plasmid was used to transfect each 6-well of mESCs, approximately 6 h after cell seeding. Transfected cells were selected with 1.0 μg ml^{-1} of puromycin 12 h later, with a change of fresh selection media every day for the next 3 days before RNA is harvested for downstream analysis. CDS of Nr5a2 and Dmrt1 were amplified from E14 mESC cDNA and cloned into the pCAG.puro vector for overexpression purposes. shRNA oligos were designed and cloned into the AgeI/EcoRI cutting site of the pLKO.1 puro vector (Addgene) according to the manufacturer's instructions.

Target sequences used for RNAi are as follows:

Luci shRNA - GATGAAATGGGTAAGTACA;

Nr5a2 shRNA - GCAAGTGTCTCAATTTAAA;

Dmrt1 shRNA1 - CGGCAGGGTTTGTTGTTATTT;

Dmrt1 shRNA2 - GCCTTCTCTGTTTCCTTACTA.

8.6. Packaging of lentivirus and establishment of stable mES cell lines

293T cells were co-transfected with pLKO.1 puro vectors or PL-SIN-EOS-C(3+)-EiP (Addgene), and viral packaging plasmids which include pCMV-VSV-G and R8.91. Transfection mix was replaced with 10% FBS-DMEM after 6 h and viral supernatant was harvested twice, once at 24 h and again at 48 h. The combined supernatant was filtered (0.45 μm) and concentrated using centrifugal filter columns (Millipore). To establish stable reporter/ knockdown cell lines, v6.4 mESCs were incubated with concentrated viruses in ES cell medium containing 8 μg ml^{-1} polybrene. After 24 h, medium was replaced with 1.0 μg ml^{-1} of puromycin selection medium. Cells were maintained/ passaged in selection medium for 1-2 weeks, before GFP-positive colonies were picked and transferred to feeder MEF for expansion and characterization as individual cell lines.

8.7. RNA extraction, reverse transcription and quantitative real-time PCR

Total RNA from cell culture was extracted using TRIzol Reagent (Invitrogen) and purified by isopropanol precipitation. Potential DNA contamination was eliminated by DNase I (Ambion) treatment at 37°C for 30 min, followed by EDTA inactivation at 65°C for 10 min. Reverse transcription was performed using SuperScript II Kit (Invitrogen). Quantitative PCR analyses were performed using a real time ABI PRISM 7900 Sequence Detection System and SYBR Green Master Mix.

8.8. Luciferase reporter assay

Selected enhancer fragments were PCR-amplified from E14 mESC genomic DNA and cloned downstream of the firefly luciferase, between the BamHI and SalI cutting

sites of a modified pGL3 basic vector. The modified pGL3 basic vector contains a 475 bp Oct4 minimal promoter fragment cloned upstream of the firefly luciferase gene. EGC or 293T cells were co-transfected with 300 ng of sequence-verified reporter construct and 0.5 ng *Renilla* luciferase construct (pRLCMV; Promega), and selected with 1.0 μg ml^{-1} of puromycin 12 h later. Firefly and *Renilla* luciferase activities were measured 48 h after transfection using the Glomax Multi-Detection System (Promega). Firefly/ *Renilla* ratios of tested enhancer regions were normalized to that of minimal promoter-only control.

8.9. EB-mediated *in vitro* differentiation

Mouse ES cells were trypsinized and cultured in non-coated 96 wells overnight in embryoid body (EB) medium (ES cell medium without β-mercaptoethanol and LIF, with 1.0 μg ml^{-1} of puromycin for selection). On day 1, the formed EBs were transferred into non-coated Petri dishes, and cultured for up to an additional 6 days. EB dissociation was performed as described previously (West et al., 2006) and GFP-positive cells were FACS-purified for RNA harvest or EGC conversion. For EGC conversion, sorted cells were cultured on feeder MEFs in mouse ES medium containing 15 ng ml^{-1} bFGF (Invitrogen), 30 ng ml^{-1} SCF (R&D Systems) and 5 μM retinoic acid (RA, Sigma). After 5-7 days of culture, alkaline phosphatase detection was performed using a commercial ESC characterization kit (Chemicon) according to the manufacturer's protocol.

9. Results and Discussion

9.1. Purification of germ cells with high stringency

It is essential to obtain pure populations of germ cells for downstream experiments. To purify mouse germ cells as material for ChIP and microarray, E13.5 gonads were dissected from *Pou5f1*-GFP transgenic mouse embryos that express GFP driven by the *Pou5f1* promoter. The *Pou5f1*-GFP reporter construct comprises 18 kb of the endogenous Oct4 genomic region but the proximal enhancer was deleted such that GFP expression is specifically activated in the ICM and germ cells, but not in the epiblast and other somatic cell types (Yeom et al., 1996). Given that germ cells exhibit sex specific differences after their arrival at the genital ridge, gonads were sexed according to the pattern of GFP expression so that male and female gonads were harvested separately. GFP-positive germ cells are organized within the testis cords in male gonads while GFP-positive germ cells are uniformly distributed in the female gonad (Figure 1a-b, d-e). Dissociation of the male and female gonads was optimized to yield single cell suspension (Figure 1c, f) for Fluorescence Activated Cell Sorting (FACS). The dissociated gonadal cells comprise a large population of GFP-negative cells and a visually distinct GFP-positive population (Figure 2). Interestingly, sex-specific differences were also observed in the GFP intensity of the GFP-positive population, with male germ cells being brighter than female germ cells. The GFP-positive fraction was routinely isolated as germ cells using a highly stringent threshold and the purity of GFP-positive cells is always in excess of 95% as indicated by post-sort analyses (Figure 2b). To validate that the *Pou5f1*-GFP reporter appropriately marks cells that express the Oct4 protein as well as to confirm that the

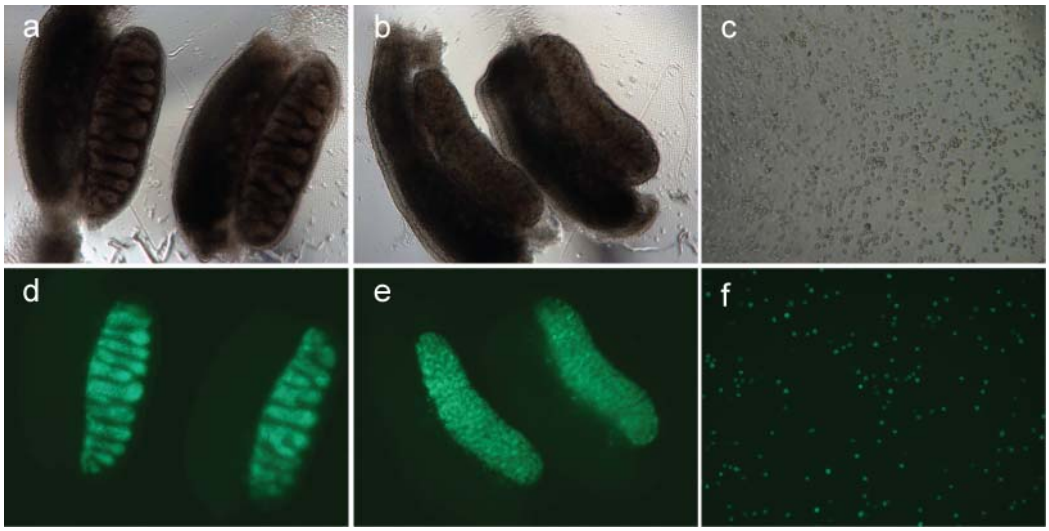

Figure 1: Isolation of gonads from *Pou5f1*-GFP transgenic mice. Phase and GFP fluorescence images of E13.5 (a,d) male gonads, (b, e) female gonads and (c, f) dissociated gonadal cells.

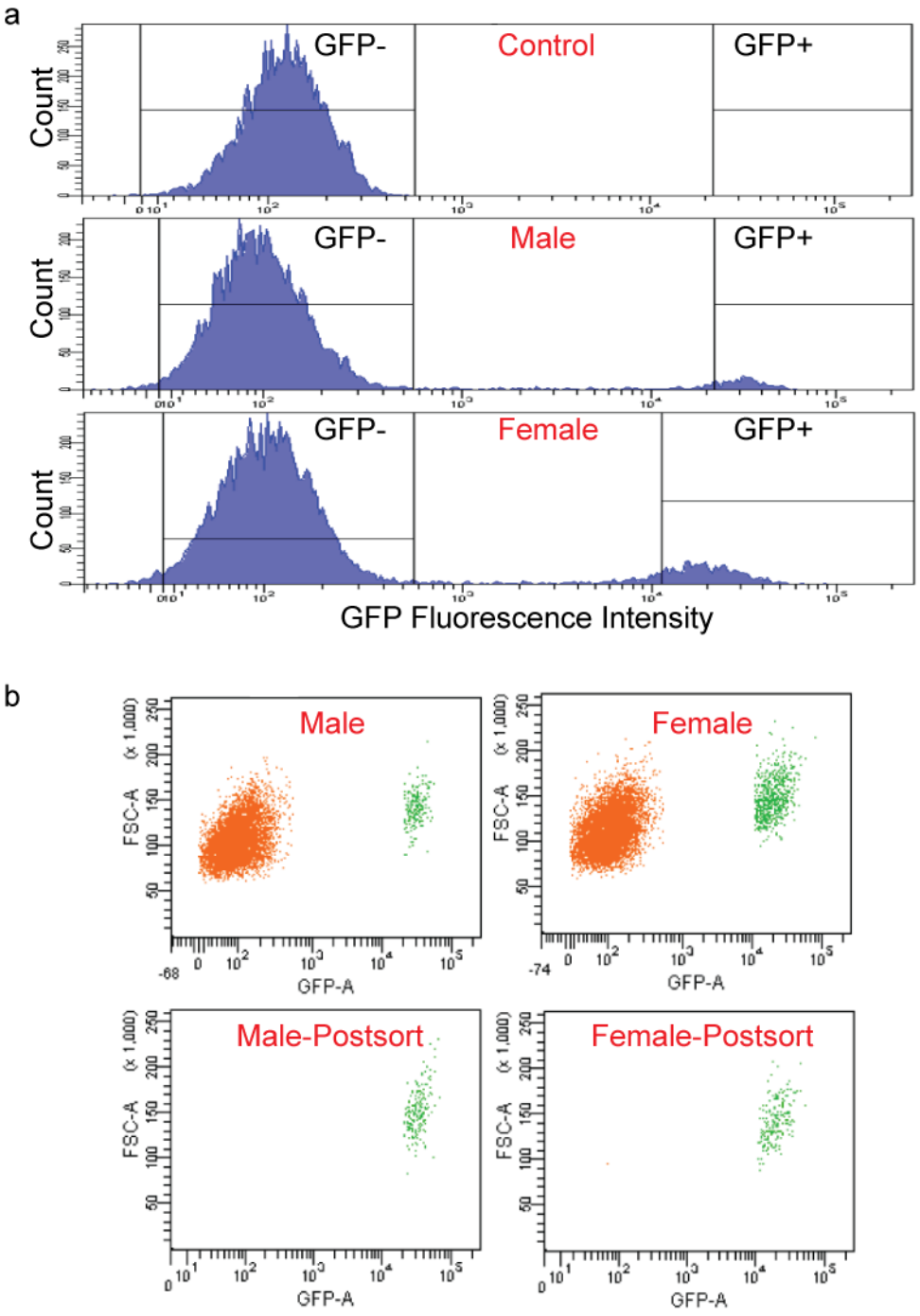

Figure 2: FACS purification of germ cells with high stringency. (a) GFP fluorescence histogram of dissociated gonadal cells isolated from E13.5 wild type control embryos and *Pou5f1*-GFP transgenic embryos. (b) Scatterplot distribution of GFP- (orange) and GFP+ (green) populations during sorting and post-sort.

post-sort results accurately reflects the efficiency of FACS purification, anti-Oct4 immunostaining was performed on the FACS- purified GFP-positive cells (Figure 3). Indeed, the GFP-positive cells also express the Oct4 protein.

9.2. Expression profiling of purified germ cells

9.2.1. Isolation of microarray-grade RNA from purified germ cells

To gain an understanding of transcriptional changes in male germ cells after their arrival at the developing gonad, germ cells were purified from E11.5, E13.5 and E15.5 embryos. In addition to counting the number of days-post-coitum, the developmental stage of embryos was also visually monitored (Figure 4). At 11.5 d.p.c., the gonad is still developing structurally and testis cords are not formed, hence the gonad cannot be sexed based on visual observation. Furthermore, the sexual identity of germ cells is only established between 11.5 to 13.5 d.p.c. (Kocer et al., 2009), depending on the signals received from the neighboring supportive somatic cells of the gonad. Hence, while male gonads were used for E13.5 and E15.5 germ cell purification, GFP-positive germ cells were harvested from E11.5 gonads regardless of the sex of the embryos.

Due to the small number of GFP-positive germ cells that can be sorted from embryos, optimization of RNA harvesting methods was essential to maximize the amount of total RNA purified. The Picopure RNA isolation kit outperforms the standard Trizol-Isopropanol precipitation method as the RNA yield was increased by more than 5 fold when only approximately 4,300 cells were used as input (Figure 5a). However, this improvement is reduced to 3 fold when cell number was increased to approximately

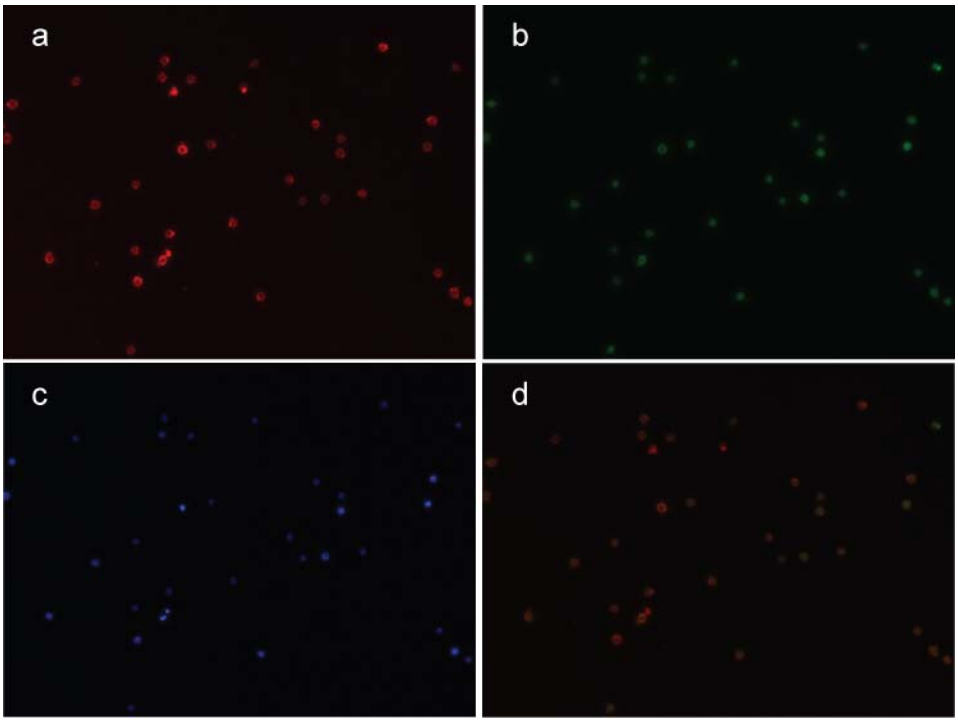

Figure 3: Validation of Oct4 expression in FACS-purified *Pou5f1*-GFP positive germ cells. (a) Oct4 immunostaining (b) *Pou5f1*-GFP (c) Hoechst (d) merged image of (a) and (b).

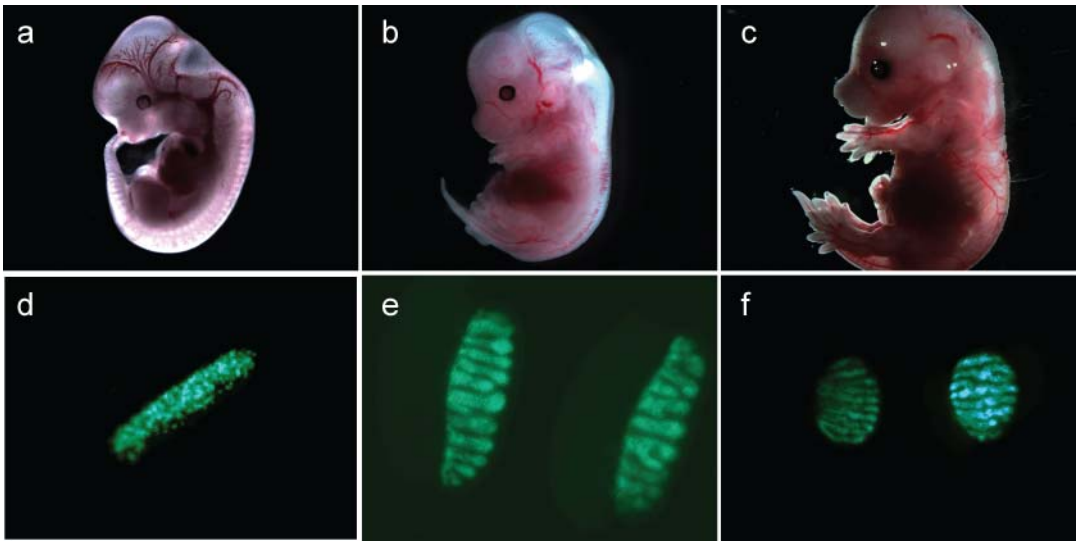

Figure 4: Development of the gonad from 11.5 d.p.c. to 15.5 d.p.c. Embryos and dissected gonads at (a,d) 11.5 d.p.c., (b,e) 13.5 d.p.c. and (c,f) 15.5 d.p.c.

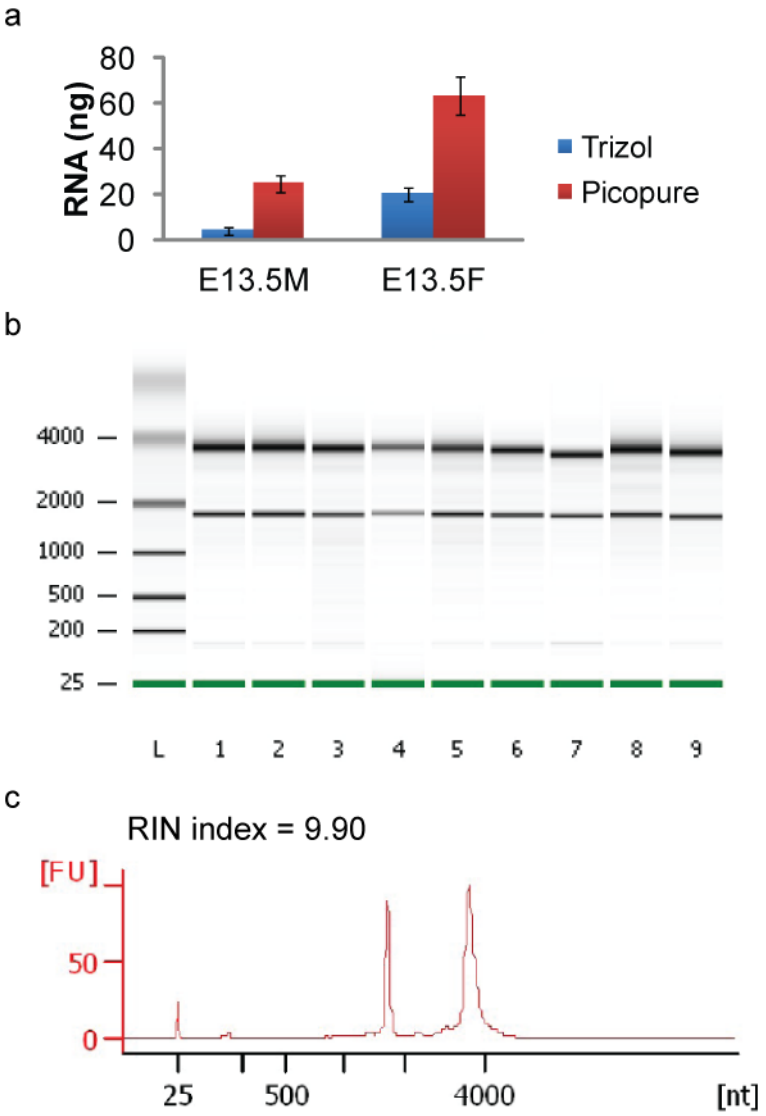

Figure 5: Harvesting microarray-grade RNA from FACS-purified germ cells. (a) Comparison of RNA yield using the Trizol method and Picopure RNA isolation kit (Arcturus) from approximately 4,300 E13.5 male germ cells (E13.5M) and 16,000 E13.5 female germ cells (E13.5F). RNA quantities were determined using the Quant-iT RiboGreen RNA Reagent. Data represents the mean ± s.e.m. of two independent experiments (n = 2). (b, c) Agilent RNA 6000 Pico electropherogram showing two distinct ribosomal bands/ peaks corresponding to 18S and 28S rRNA. A RIN index greater than 8 also indicates good quality RNA.

16,000 cells. Overall, the Picopure RNA isolation kit is superior to the Trizol method when RNA is isolated from small cell populations. As estimation, it is possible to isolate more than 20 ng of total RNA from less than 5,000 sorted germ cells.

The quality of isolated RNA was examined using the RNA 6000 Pico LabChip Kit. The two distinct bands on the gel electropherogram (Figure 5b,c) which represents the 18S and 28S rRNA, and a uniformly low baseline, indicates good quality RNA with little degradation and the absence of DNA contamination. The RNA Integrity Number (RIN) is a number system that is commonly used to grade RNA quality, with 1 being the most degraded and 10 being the most intact or high quality RNA. Generally, it is recommended to use samples that have RIN greater than 8.0 for microarray. Since RNA quality is crucial for achieving success in microarray experiments, only samples that have RIN greater than 9.0 were processed for downstream experiments.

9.2.2. Microarray analysis and qPCR validation

Despite optimized RNA isolation protocol from FACS-purified germ cells, the amount of RNA obtained is much less than 500 ng, which is the recommended starting RNA quantity for the *in vitro* transcription (IVT) kit from Ambion. Therefore, I have used the Ovation RNA amplification strategy followed by *in vitro* biotin-labeling (Illustration 4). Microarray was performed on E11.5, E13.5 and E15.5 germ cells, and E13.5 GFP-negative non-germ cells from the gonad were used as negative control. Hierarchical clustering of the microarray samples using all (> 45 200) probes reveals that E11.5, E13.5 and E15.5 germ cells cluster together and differ from the non-germ cell cluster (Figure 6). All biological replicates (harvested and purified on independent days) of the same time point cluster together, rendering confidence to the

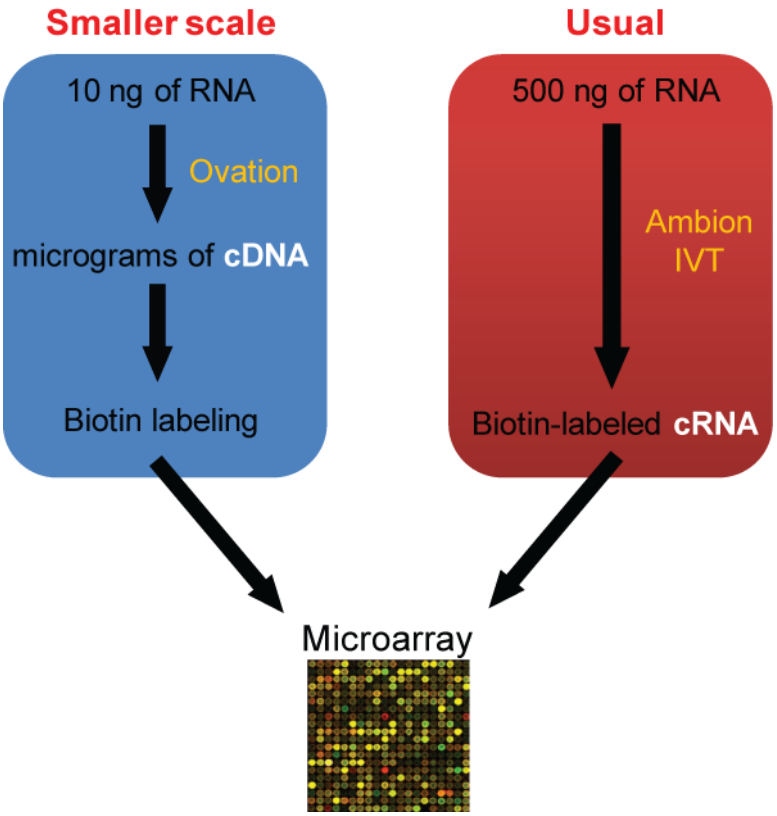

Illustration 4: Microarray strategy using small quantities of RNA. 10 ng of RNA isolated from FACS purified germ cells was amplified using the OVATION RNA amplification V2 kit (NuGEN), after which the cDNA was biotin-labeled and hybridized to the Mouse WG6-v2 microarrays (Illumina).

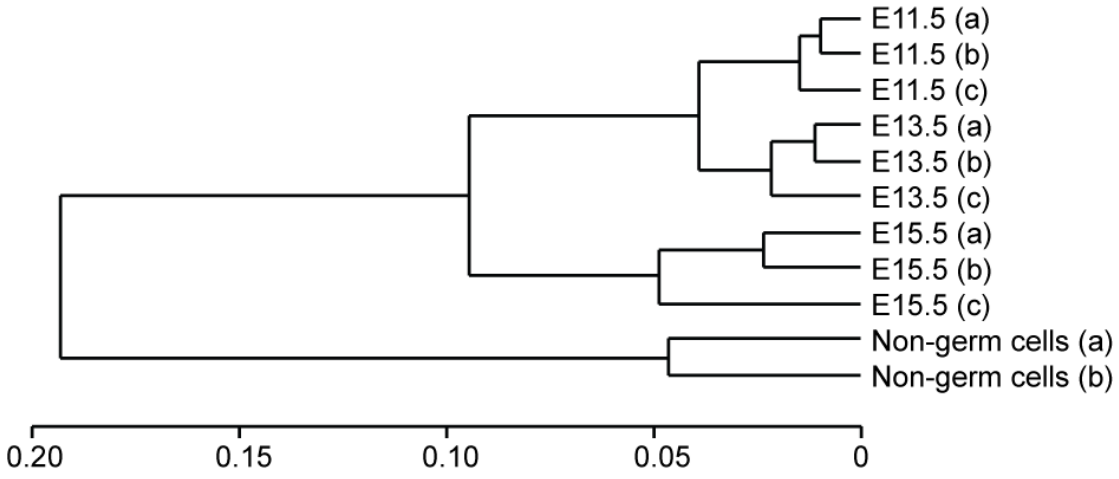

Figure 6: Hierarchical clustering of microarray samples. Biological triplicate data (a,b and c) are shown for E11.5, E13.5 and E15.5 germ cells while biological duplicate data (a and b) are shown for Non-germ cells as control. The x-axis represents the extent of dissimilarity between nodes of the dendrogram.

data quality. Notably, the clustering result also suggests that E13.5 germ cells are more closely related to E11.5 germ cells as compared to E15.5 germ cells, at least on a transcriptome level.

Closer inspection of specific transcripts reveals that pluripotency-associated markers such as *Sox2*, *Nanog* and *Prdm14* are downregulated from E11.5 to E15.5 (Figure 7a). On the other hand, *Pou5f1* is only marginally downregulated and still remains enriched in E15.5 germ cells compared to non-germ cells. Post-migratory germ cell markers *Dazl* and *Ddx4* are also upregulated by E13.5. While *Dazl* is already highly expressed at E11.5, it peaks in expression at E13.5. *Ddx4* shows a consistent upregulation from E11.5 to E13.5 and its high expression is maintained at E15.5. Realtime quantitative PCR (qPCR) was performed to validate the microarray data (Figure 7b) and shows similar trends.

It has previously been shown that the potential of *in vivo*- derived germ cells to be converted to pluripotent EGC declines from E8.5 to E12.5 (Labosky et al., 1994). One possible explanation is the downregulation of pluripotency-associated genes in post-migratory germ cells when they arrive at the developing gonad. By comparing microarray data of mESCs (v6.4 mESC and R1 mESC) and MEFs (*Pou5f1*-GFP MEF and *Actin*-GFP MEF) (Heng et al., 2010), an ESC-enriched gene list was generated to include the top 100 genes that are at least 5-fold upregulated in ESCs compared to MEFs (q-value < 0.01). A microarray heatmap was then plotted to illustrate the transcriptional changes of these ESC-enriched genes in germ cells (Figure 8). Two-class unpaired T-test was performed to compare germ cells of each time point to non-

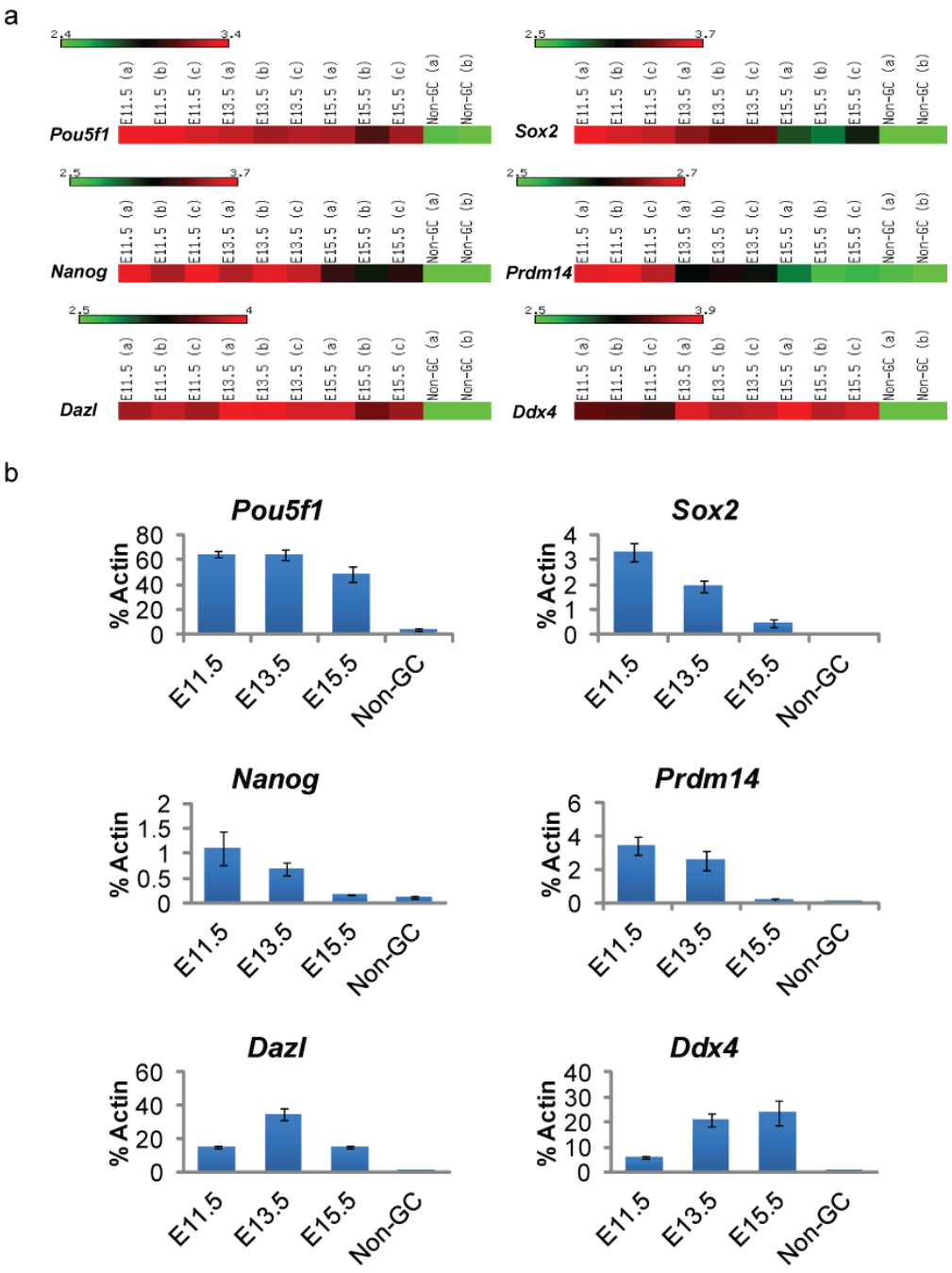

Figure 7: qPCR validation of microarray data. (a) Microarray heatmaps (log-transformed, mean-centered) showing biological triplicate data (a,b and c) for E11.5, E13.5 and E15.5 germ cells and biological duplicate data (a and b) for Non-germ cells (Non-GC) as control. (b) qPCR validation for selected pluripotency and germ cell markers. Data represents the mean ± s.e.m. of three biological replicates (n = 3).

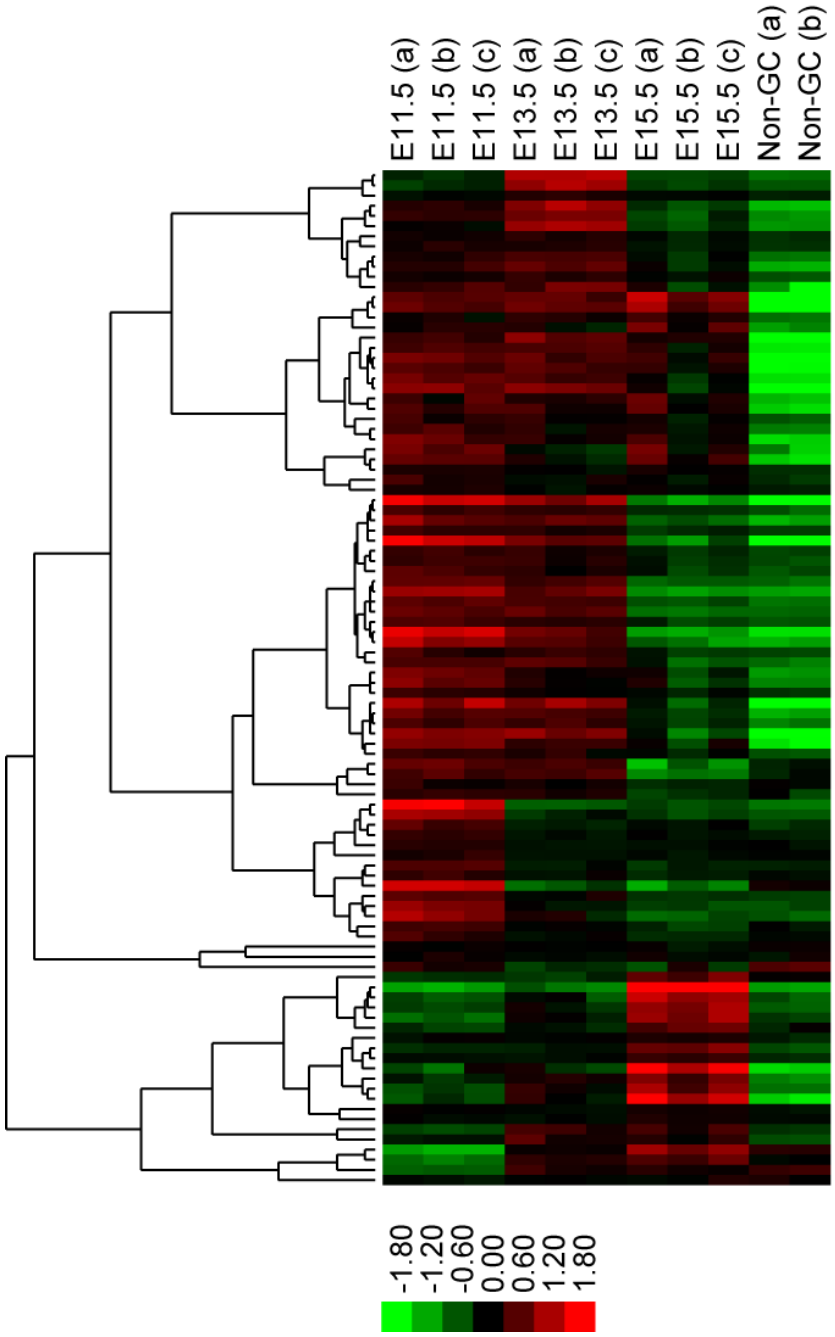

Figure 8: Microarray heatmap illustrating expression changes of top 100 ESC-enriched genes. Data was log-transformed and mean-centered prior to plotting the heatmap. Biological triplicate data (a,b and c) are shown for E11.5, E13.5 and E15.5 germ cells while biological duplicate data (a and b) are shown for Non-germ cells (Non-GC) as control.

germ cells. Using a cut-off of fold change > 1.5 and q-value < 0.01, 46% of the ESC-enriched genes are upregulated in E11.5 germ cells, 48% in E13.5 germ cells and 37% in E15.5 germ cells. In combination, 74% of ESC-enriched genes are more highly expressed in germ cells (E11.5, E13.5 and/or E15.5) than non-germ cells.

This group of genes that are upregulated in both ESC and germ cells is enriched for Gene Ontology (GO) categories such as transcription factor (p-value = 5.1×10^{-5}), stem cell maintenance (p-value = 8.0×10^{-7}) and transforming growth factor-beta (TGFβ)- related genes (p-value = 4.9×10^{-3}). These genes also show variable patterns of expression changes as germ cells progress from E11.5 to E15.5 (Figure 8). As expected, many ESC-associated genes are downregulated during this period (Figure 9a). These downregulated ESC-associated genes such as *Pou5f1*, *Sox2*, *Nanog*, *Esrrb*, *Dppa2*, *Dppa4*, *Klf2*, *Klf5*, *Zic3*, *Rfx2*, *Utf1*, *Nr5a2*, *Lin28*, *Cdh1* and *Tcfcp2l1* have been shown to be important in the maintenance of pluripotency in ESCs or reacquisition of pluripotency during somatic cell reprogramming (Feng et al., 2009a; Jiang et al., 2008; Kooistra et al., 2010; Lim et al., 2007; Pardo et al., 2010; Redmer et al., 2011; Takahashi and Yamanaka, 2006; van den Berg et al., 2010; Yu et al., 2007; Zhang et al., 2006). Curiously, genes related to the TGFβ signaling pathway, *Gdf3*, *Lefty1* and *Nodal*, show elevated expression in E13.5 germ cells followed by downregulation at E15.5 (Figure 9b). This may suggest an importance of TGFβ signaling pathway in E13.5 germ cells, which is consistent with a previous report that TGFβ signaling pathway negatively regulates fetal and postnatal germ cell proliferation (Moreno et al., 2010). On the other hand, some ESC-enriched genes are upregulated instead. For example, the upregulation of *Dnmt3a* and *Dnmt3l* (Figure 9c)

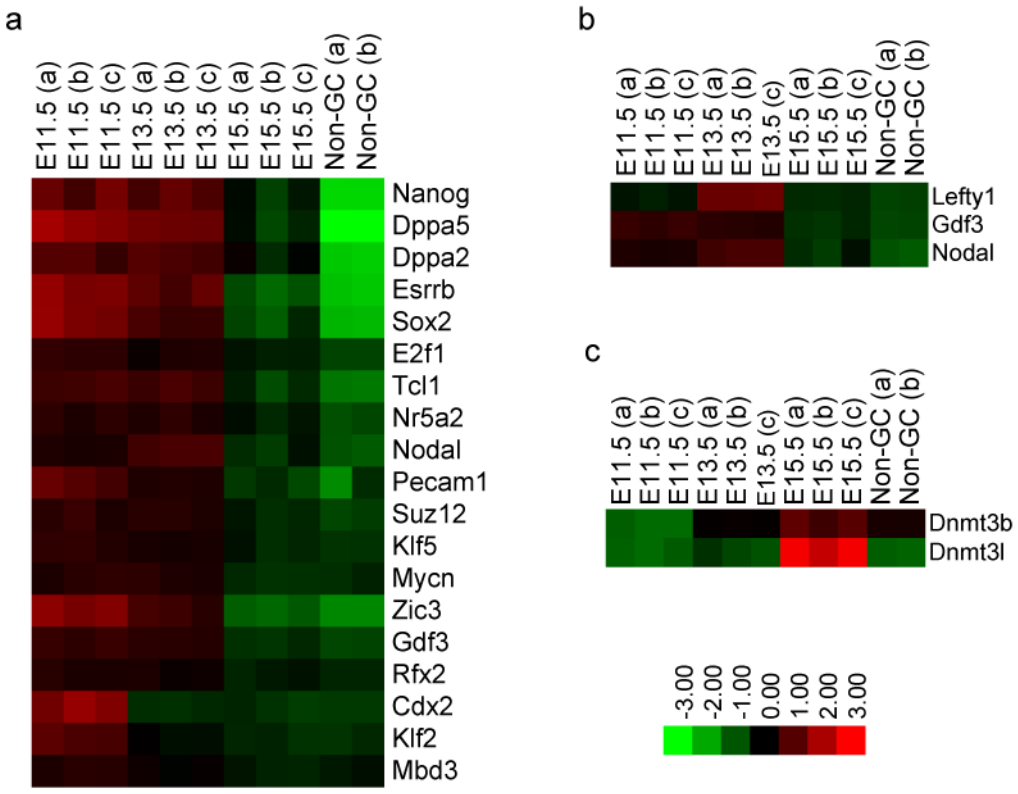

Figure 9: Microarray heatmaps of downregulated and upregulated ESC-enriched genes. (a) Pluripotency-associated genes (b) Tgfβ signaling pathway members (c) Dnmt3 and Dnmt3l. Biological triplicate data (a,b and c) are shown for E11.5, E13.5 and E15.5 germ cells while biological duplicate data (a and b) are shown for Non-germ cells (Non-GC) as control.

may be crucial for male germ cell imprinting (Bourc'his et al., 2001; Hata et al., 2002).

To gain further insight into major biological changes that occur from E11.5 to E15.5, genes that exhibit > 1.5-fold downregulation from E11.5 to E15.5, and are > 1.5-fold enriched in E11.5 compared to non-germ cells were further analyzed. Enriched GO annotation clusters for biological processes include cell cycle (p-value = 7.71×10^{-12}), DNA packaging (p-value = 5.59×10^{-5}), stem cell maintenance (p-value = 4.90×10^{-4}), RNA splicing (p-value = 3.21×10^{-4}) and germ cell migration/ apoptosis (p-value = 4.76×10^{-3}) (Figure 10a). Transcription factor is also enriched as GO category for molecular function (p-value = 6.87×10^{-3}). The downregulation of cell cycle genes such as *Ccnb1*, *Cdca3* and *c-Myb* indicate reduced mitotic entry after E13.5 (Ayad et al., 2003; Nakata et al., 2007). DNA packaging genes include histone genes (*Hist1h2af*, *Hist1h2ad*, *Hist1h1a*, *Hist2h3c1*, *Hist2h2ac*, *Hist1h2bj*, *Hist1h2ai*, *Hist1h2ah*, *Hist1h3e*, *Hist1h2ao*) and *HELLS*, which is a member of the SNF2 family of chromatin remodeling proteins shown to modulate DNA methylation (Tao et al., 2011), spermatogonial proliferation and meiotic progression in germ cells (De La Fuente et al., 2006; Zeng et al., 2011). Therefore, the microarray data suggests that molecular changes that occur involve multiple levels of regulation such as epigenetics, transcriptional control and even post-transcriptional splicing. Visual inspection of the microarray heatmaps plotted for downregulated genes also reveals that majority of stem cell maintenance genes and cell cycle genes are transcriptionally downregulated after E13.5 (Figure 9a, 11b-c). On the other hand, germ cell migration genes and apoptosis-related genes are downregulated earlier and more gradually from E11.5 to E15.5 (Figure 11a).

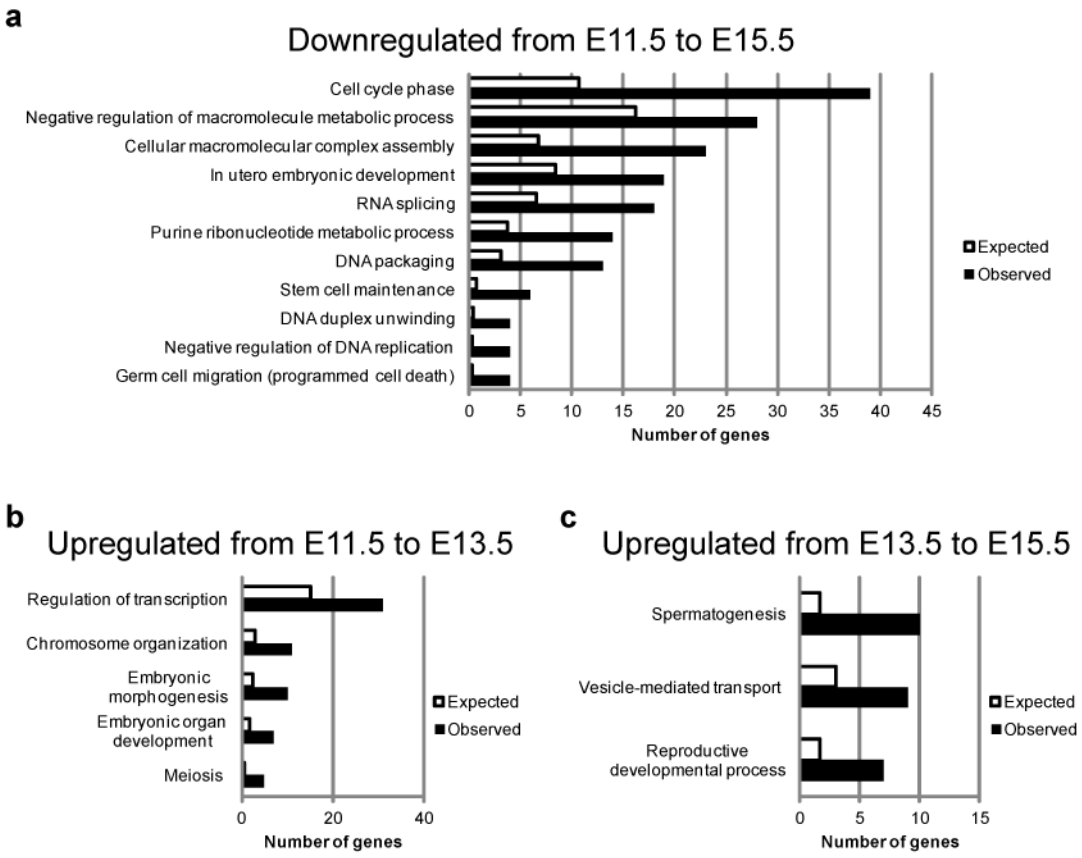

Figure 10: Enriched GO biological process categories for upregulated and downregulated genes. Significant GO categories ($p < 0.05$) for genes that are (a) downregulated from E11.5 to E15.5, (b) upregulated from E11.5 to E13.5 or (c) upregulated from E11.5 to E15.5.

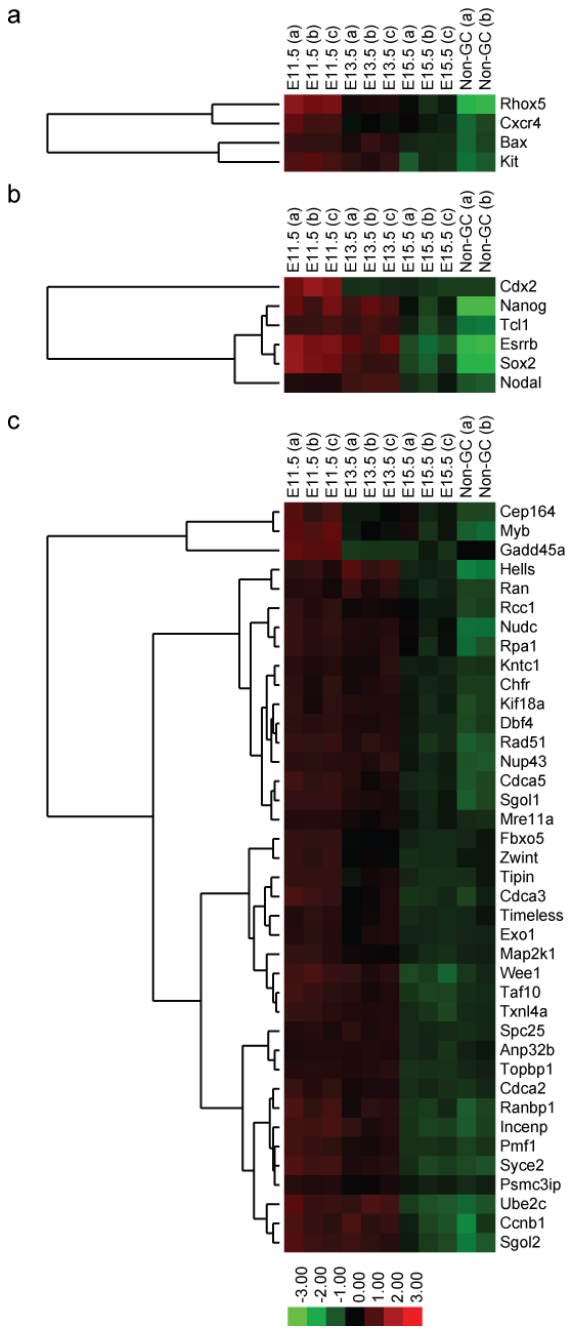

Figure 11: Microarray heatmaps of downregulated genes (from E11.5 to E15.5) related to (a) germ cell migration and programmed cell death (b) stem cell maintenance and (c) cell cycle. Biological triplicate data (a,b and c) are shown for E11.5, E13.5 and E15.5 germ cells while biological duplicate data (a and b) are shown for Non-germ cells (Non-GC) as control.

Upregulated genes were similarly analyzed. Enriched GO categories for genes that are > 1.5-fold upregulated from E11.5 to E13.5, and at least 1.5-fold enriched in E13.5 compared to non-germ cells were examined (Figure 10b). The top significant GO category for biological process is related to regulation of transcription (p-value = 6.94 x 10^{-5}) (Figure 12a), which is defined as "any process that modulates the frequency, rate or extent of cellular DNA-dependent transcription". Transcription factors also form an enriched GO category for molecular function (p-value = 1.05 x 10^{-3}). One notable upregulated transcription factor is *Cdkn2a*, also known as $p16^{INK4A}$, which arrests cell cycle in G1 phase by its interaction with CDK4. Importantly, *Cdkn2a* poses a barrier to iPS reprogramming, and its inhibition improves reprogramming efficiency (Li et al., 2009). The second most significant GO category for biological process is chromosome organization (p-value = 4.04 x 10^{-4}) (Figure 12b) which includes the chromatin remodeling gene *Arid4a* with functions in genomic imprinting (Wu et al., 2006a) and cell cycle arrest (Lai et al., 2001), NuRD complex subunit *Rbbp7*, as well as H3K4 methyltransferases *Setd7* and *MLL3*.

Generally, genes that are upregulated at E13.5 may be clustered into two groups (Figure 12). Group 1 includes genes that are upregulated at E13.5 and maintained at E15.5, while group 2 includes genes that are upregulated transiently at E13.5. Group 1 genes, such as *Cdkn2a*, may explain why E13.5 germ cells cannot be reverted to pluripotent EGC, despite the observation that many of the pluripotency-associated genes have not been fully repressed. Group 2 genes, on the other hand, may be required transiently for the germ cells to make a switch to commit towards germ cell differentiation. More experiments would be required to test these hypotheses.

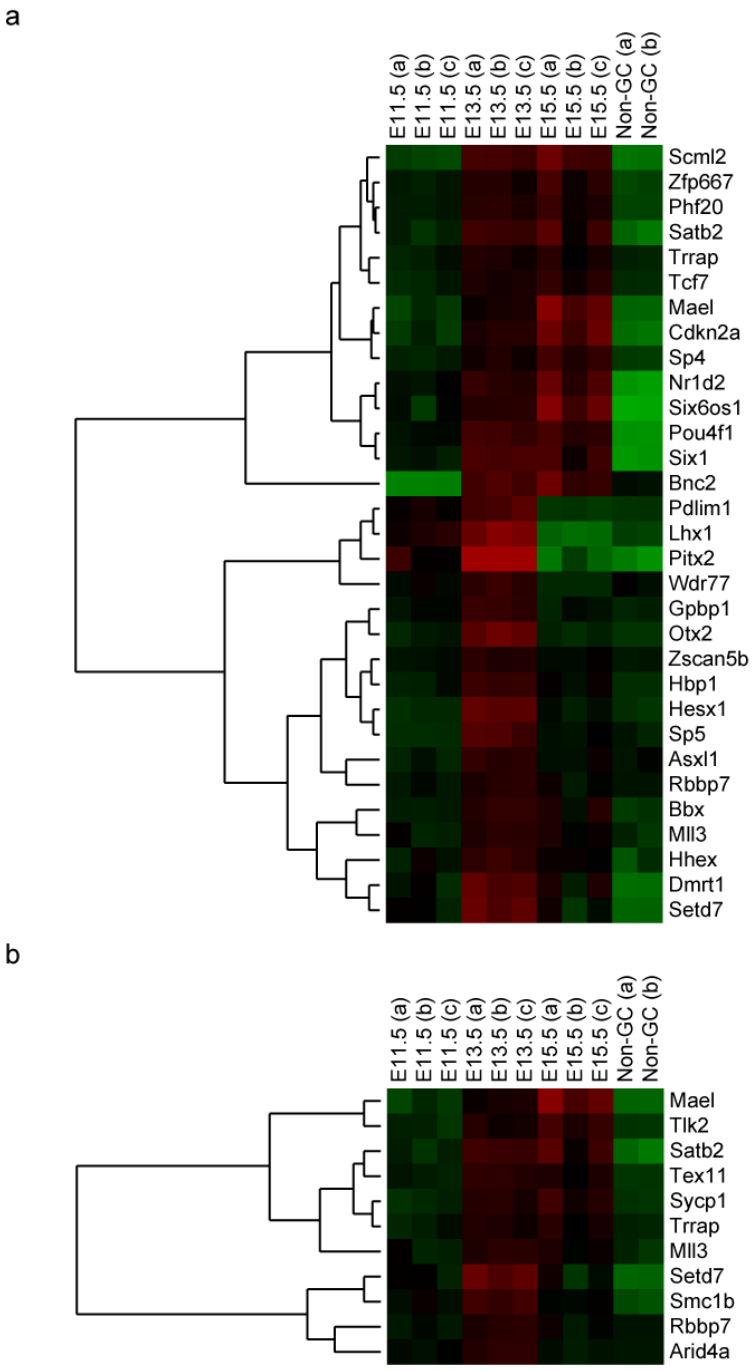

Figure 12: Microarray heatmaps of upregulated genes (from E11.5 to E13.5) related to (a) regulation of transcription and (b) chromosome organization. Biological triplicate data (a,b and c) are shown for E11.5, E13.5 and E15.5 germ cells while biological duplicate data (a and b) are shown for Non-germ cells (Non-GC) as control.

Lastly, genes that are > 1.5-fold upregulated from E11.5 to E15.5, and at least 1.5-fold enriched in E15.5 compared to non-GC are enriched for spermatogenesis (p-value = 2.98 x 10^{-5}) and reproductive developmental process (p-value = 6.18 x 10^{-3}) (Figure 10c). These genes are largely upregulated later between E13.5 to E15.5 (Figure 13). One of the earliest upregulated spermatogenesis-related genes is *Nanos2* which is upregulated more than 9-fold at E13.5. *Nanos2* encodes for the RNA-binding protein that has been previously shown to prevent meiosis and activate the male germ cell differentiation program (Suzuki and Saga, 2008).

In summary, the microarray data analyses suggest that germ cells undergo major transcriptional changes between E11.5 to E15.5, especially after E13.5. Many ESC-enriched genes are also enriched in germ cells compared to non-germ cells, including stem cell maintenance genes that are largely downregulated between E13.5 to E15.5. Most of the genes that promote mitotic progression are also transcriptionally repressed after E13.5. Contrary to repression of the downregulated genes, induction of some genes occurs earlier between E11.5 to E13.5. Notable examples include *Cdkn2a* and *Nanos2* that are likely to reduce germ cell "pluripotential" and promote expression of other germ cell differentiation genes (Illustration 5). Lastly, majority of genes involved in spermatogenesis are only upregulated later at E15.5. In addition, GO analyses suggest that these molecular changes involve both transcription factors and chromosomal organization, hence providing motivation to study epigenetic properties of germ cells during this dynamic period. Moreover, germ cells at 13.5 d.p.c. are undergoing a transitional phase and provide a good timepoint to study events that trigger the loss of "pluripotential" and activation of germ cell differentiation.

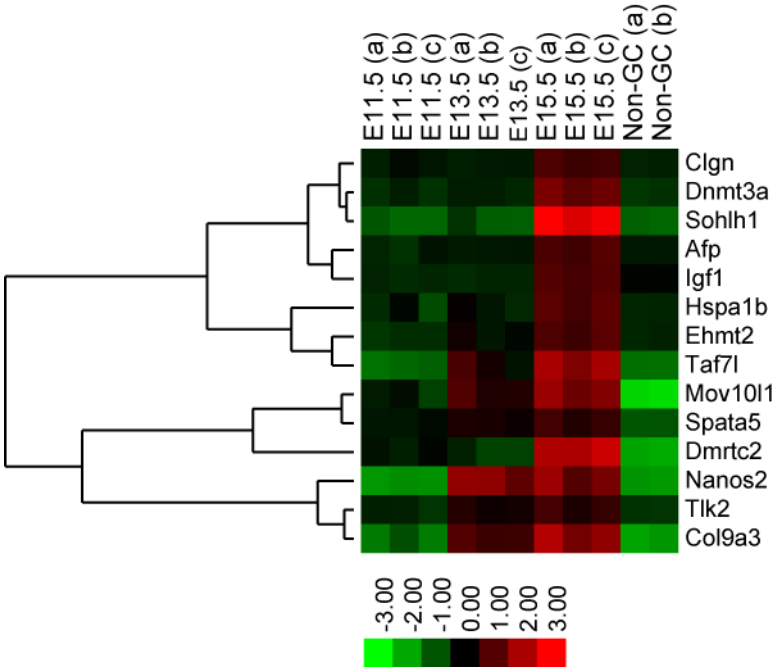

Figure 13: Microarray heatmap of upregulated genes (from E11.5 to E15.5) related to spermatogenesis and reproductive processes. Biological triplicate data (a,b and c) are shown for E11.5, E13.5 and E15.5 germ cells while biological duplicate data (a and b) are shown for Non-germ cells (Non-GC) as control.

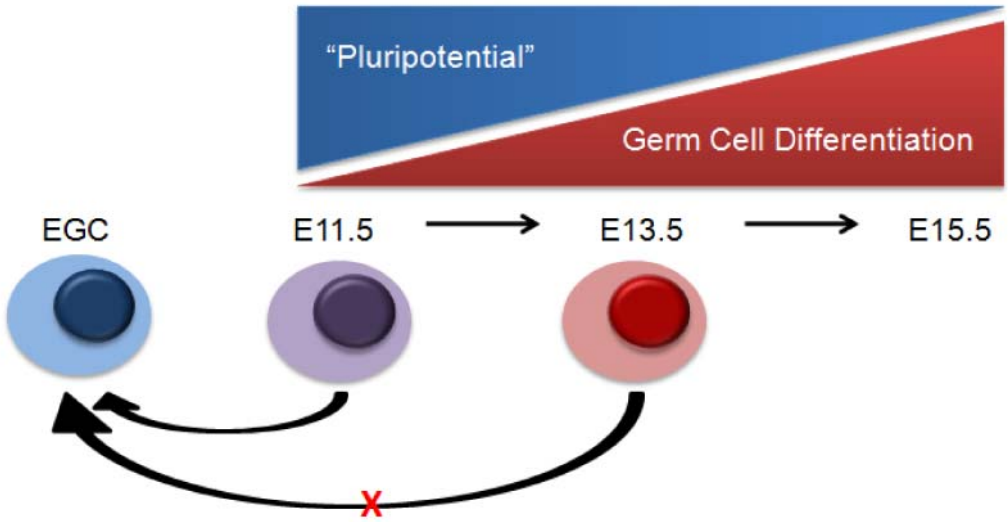

Illustration 5: Developmental progression of germ cells from E11.5 to E15.5. From E11.5 to E15.5, male germ cells lose pluripotential and the ability to revert to EGC while undergoing male germ cell differentiation.

9.3. Small-scale ChIP on purified germ cells

9.3.1. Optimization of small-scale ChIP using mESCs

Before embarking on the actual germ cell ChIP, I first optimized the small-scale ChIP procedures using mESCs that are abundantly available from cell cultures. Since the ultimate goal is to generate genome-wide ChIP-Seq libraries from small numbers of FACS-purified cells, an amplification step is essential for providing sufficient DNA for library preparations. Our laboratory (Dr Muratani Masafumi) has established a small-scale ChIP procedure (Illustration 6), whereby de-crosslinked ChIP DNA is subjected to Whole Genome Amplification (WGA) followed by two rounds of BpmI restriction digestion to excise the PCR adaptors prior to ChIP-Seq library preparation.

Experiments were conducted to examine the effects of WGA on ChIP. A normal-scale ChIP was performed against H3K4me3 using mESC chromatin, and small amounts of de-crosslinked DNA (50 pg, 250 pg and 500 pg) as determined by Picogreen quantitation were subjected to 15 cycles of WGA. As compared to control regions, ChIP enrichment was detected at the *Pou5f1* promoter and known bivalent promoters such as *Zfpm2* and *Dlx1* both before and after 15 cycles of WGA (Figure 14a,b). The pattern of relative enrichment across different tested regions appears unaffected by WGA, but a reduction in fold enrichment was observed when starting amount of DNA is reduced to 50 pg. Nevertheless, 15 cycles of WGA from 50 pg of de-crosslinked ChIP DNA provides a yield of about 500 ng, indicating an amplification of approximately 10,000 fold (Figure 14c). Amplification has also not reached saturation

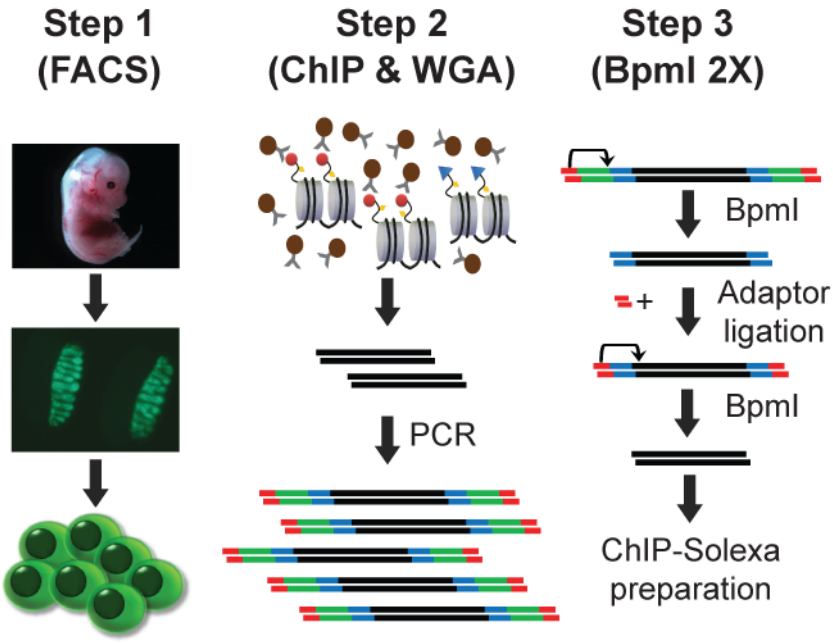

Illustration 6: Experimental workflow of germ cell ChIP-Seq. Germ cells were FACS-purified based on *Pou5f1-GFP* expression. Following ChIP, the de-crosslinked immunoprecipitated DNA was amplified by WGA. Amplification adaptors were cleaved prior to ChIP-Seq library preparation.

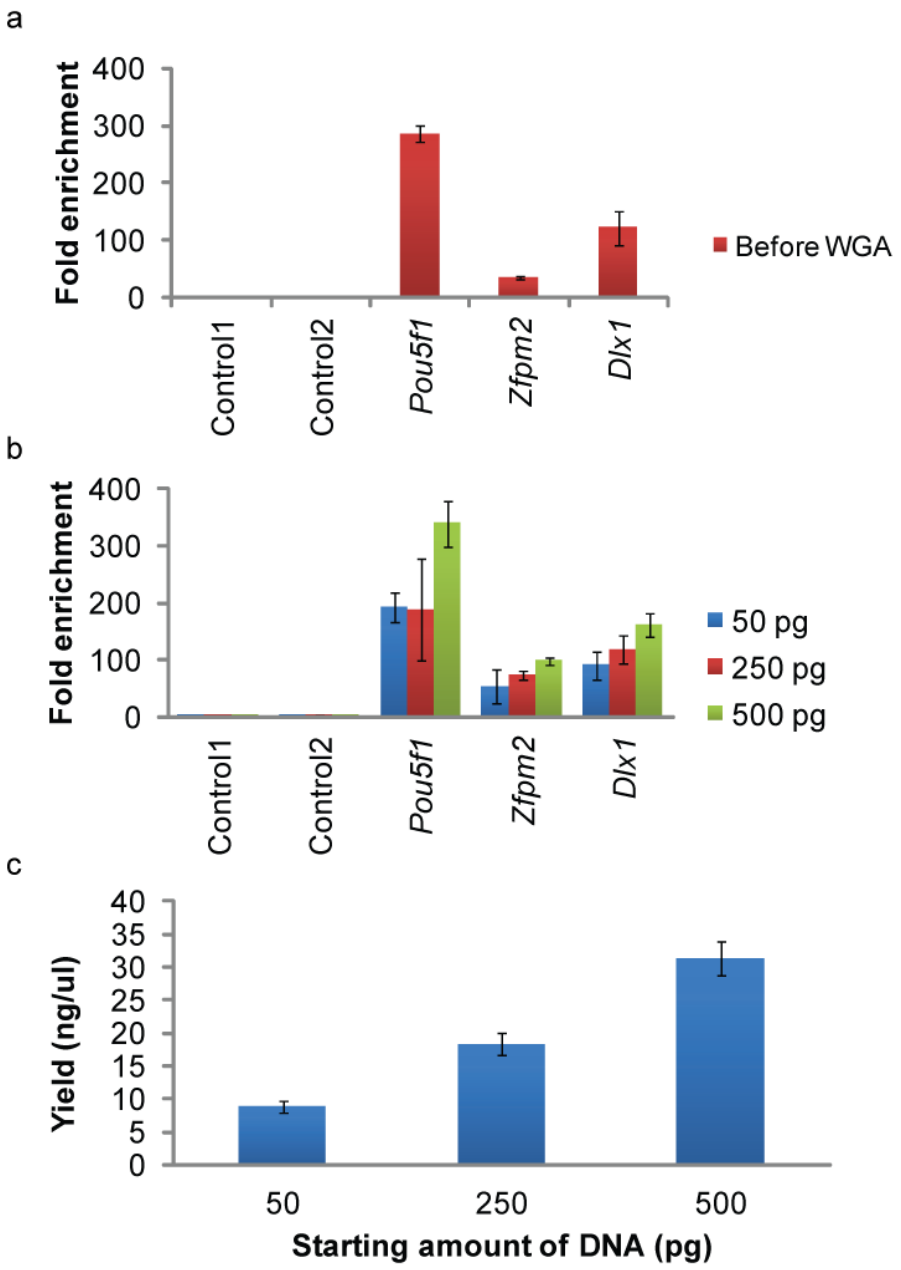

Figure 14: Similarity of H3K4me3 ChIP signals before and after WGA amplification. (a) qPCR results of normal-scale H3K4me3 ChIP performed on E14 mESC (a) without amplification and (b) after 15 cycles of WGA amplification from 50 pg, 250 pg or 500 pg of ChIP DNA. (c) DNA yield after 15 cycles of WGA amplification. Data represents the mean ± s.e.m. of three independent experiments (n = 3).

with 15 WGA cycles, as DNA yield continues to increase with amount of input DNA (Figure 14c). ChIP-Seq libraries were generated using the samples that were WGA-amplified from 50 pg of de-crosslinked H3K4me3 ChIP DNA. The WGA technical duplicates (WGA1 and WGA2) show similar peaks to each other and the non-amplified library (Figure 15a). On a genome-wide level, WGA1 and WGA2 are highly correlated (r = 0.977) and have a peak overlap of 95.6%. When WGA1 or WGA2 is compared to the non-WGA library, correlation remains high (r = 0.898 and 0.896) with peak overlap of 86.4% and 86.1% respectively (Figure 15b). In summary, WGA provides an attractive amplification tool for small-scale ChIP.

Next, I proceeded to perform the small-scale ChIP procedures on small numbers of mESCs. Compared to a typical ChIP, procedures have been fine-tuned to reduce material loss and background levels. Small-scale H3K4me3 ChIP gave good fold enrichments at tested sites when 100,000 mESCs were used without WGA amplification (Figure 16a). It is difficult to examine ChIP enrichment by qPCR if the starting cell numbers were reduced to below 100,000 due to inherent qPCR detection limitations. Therefore, all subsequent experiments included the WGA step prior to qPCR. ChIP enrichment could be achieved from as little as 10,000 mESCs, but the fold enrichment decreased as cell numbers were reduced from 100,000 to 10,000 (Figure 16b). To achieve good quality ChIP-Seq, it is always better to begin with larger cell numbers, especially given the consideration that H3K4me3 antibody provides one of the best enrichments among the tested antibodies.

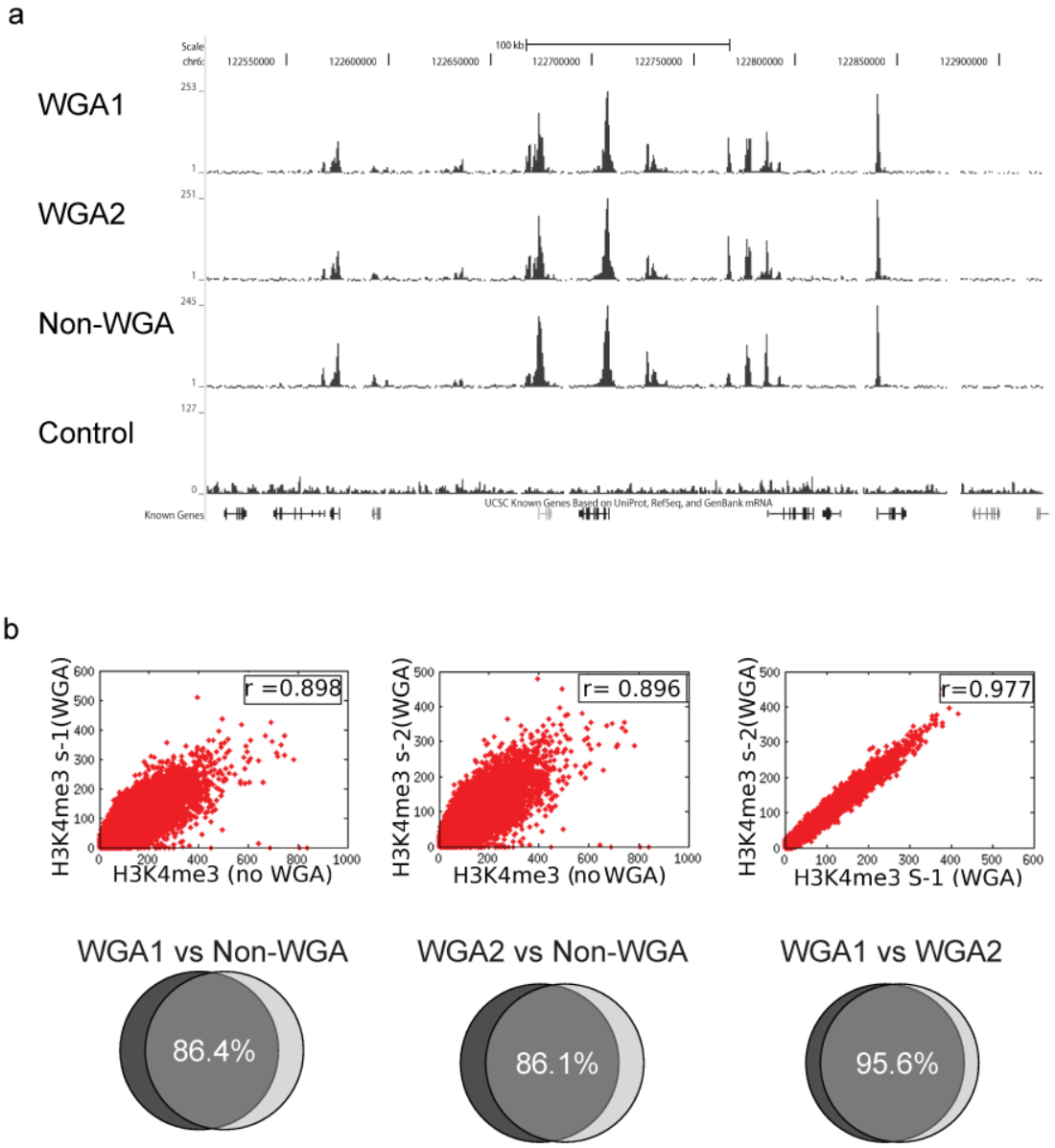

Figure 15: Genome-wide comparisons of H3K4me3 ChIP before and after WGA amplification. (a) UCSC browser view of H3K4me3 ChIP-Seq signals. (b) Scatterplot of H3K4me3 ChIP-Seq signals and Venn diagram illustrating overlap of peaks between amplified (WGA1 and WGA2) and non-amplified samples (Non-WGA).

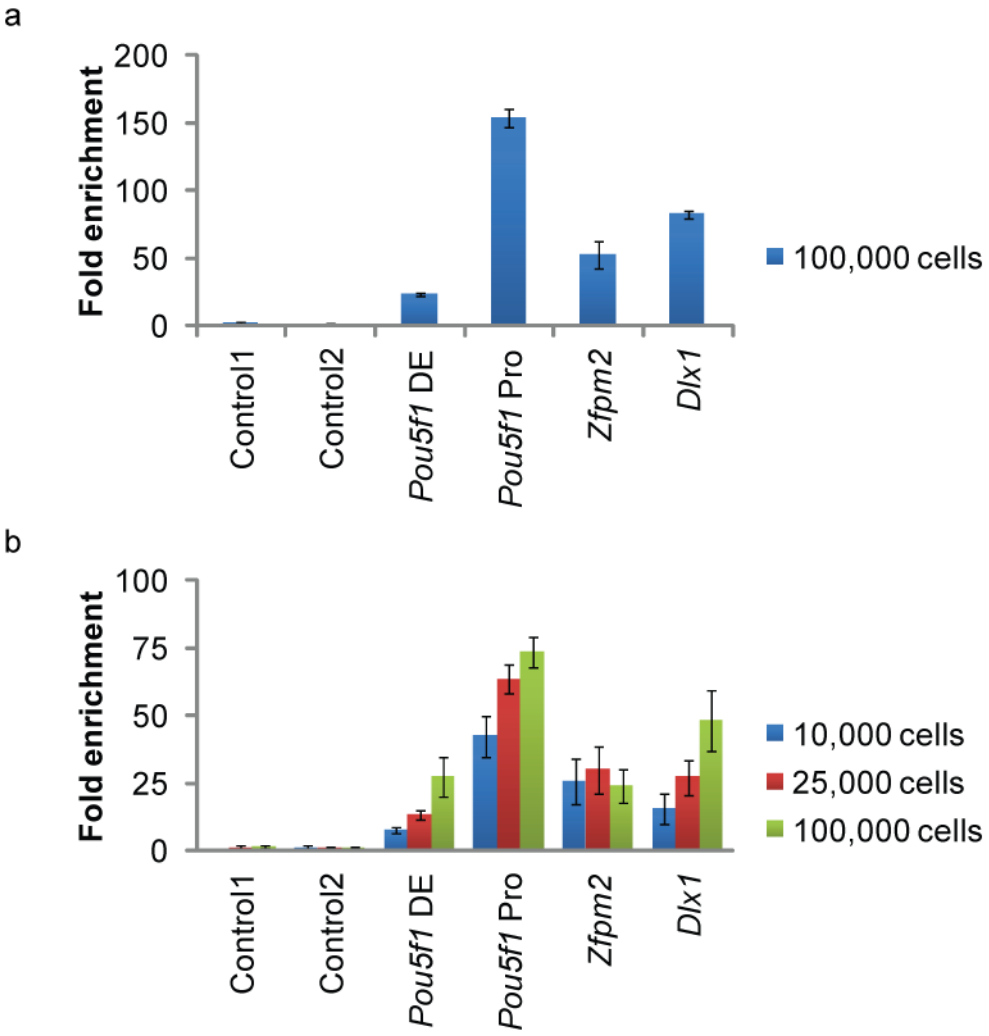

Figure 16: Small-scale ChIP could be performed on fewer numbers of cells. (a) qPCR results of small-scale H3K4me3 ChIP performed on 100,000 E14 mESC without WGA amplification. (b) qPCR results of small-scale H3K4me3 ChIP performed on 10,000, 25,000 and 100,000 E14 mESC, after 15-18 cycles of WGA amplification. *Pou5f1* DE and *Pou5f1* Pro refer to the distal enhancer and proximal promoter of the *Pou5f1* gene respectively. Data represents the mean ± s.e.m. of three independent experiments (n = 3).

9.3.2. Germ cell ChIP-Seq led to identification of active regulatory elements

GFP-positive germ cells were purified, fixed and frozen at -80°C until sufficient cell numbers are achieved (Illustration 6, Step 1). Cells were then pooled so that approximately 50,000 to 100,000 germ cells were used for each ChIP experiment. ChIP-Seq libraries were generated in duplicates for H3K4me3, H3K27me3, H3K27ac and H2BK20ac after enrichment was validated by qPCR (Figure 17).

Reproducibility was examined by plotting tag counts across 1-kb bins. The ChIP duplicates for H3K4me3 and H3K27me are highly correlated (r = 0.87 and 0.85) (Figure 18a), and visual inspection of signals in the UCSC browser also shows similar peaks at gene promoters in the H3K4me3 duplicates (Figure 18b). However, the duplicates for H3K27ac and H2BK20ac are less well-correlated (r = 0.74 and 0.75). Stringent criteria were thus set for downstream analyses: (1) peak detection using p-value < 0.05 and (2) only peaks that overlap in biological duplicates are used.

To investigate how the occurrence of histone marks at promoter regions correlate with expression levels, genes were first grouped into 3 categories based on microarray expression: (1) high = above 80^{th} percentile, (2) medium = 40-60^{th} percentile and (3) low = below 30^{th} percentile. The histone ChIP-Seq signals were plotted for each of these 3 categories, spanning 4 kb around known transcriptional start sites (TSS) (Figure 19). H3K4me3 and H3K27ac positively correlate with expression levels, while H3K27me3 show an inverse relationship. This observation supports the notion that H3K4me3 and H3K27ac mark active gene promoters and H3K27me3 mark poised/ repressed promoters. The H3K27me3 plots had a broad shape that is typical

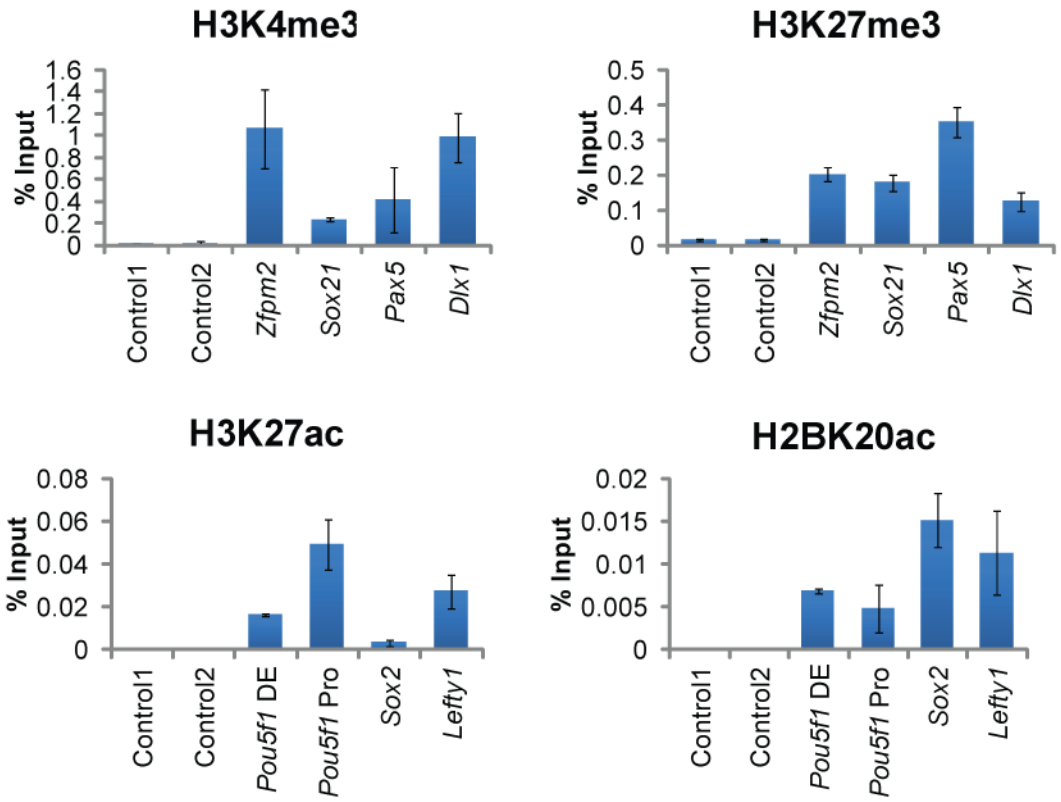

Figure 17: Histone ChIP-qPCR results of E13.5 male germ cells. Data represents the mean ± s.e.m. of two independent experiments (n = 2). *Pou5f1* DE and *Pou5f1* Pro refer to the distal enhancer and proximal promoter of the *Pou5f1* gene respectively.

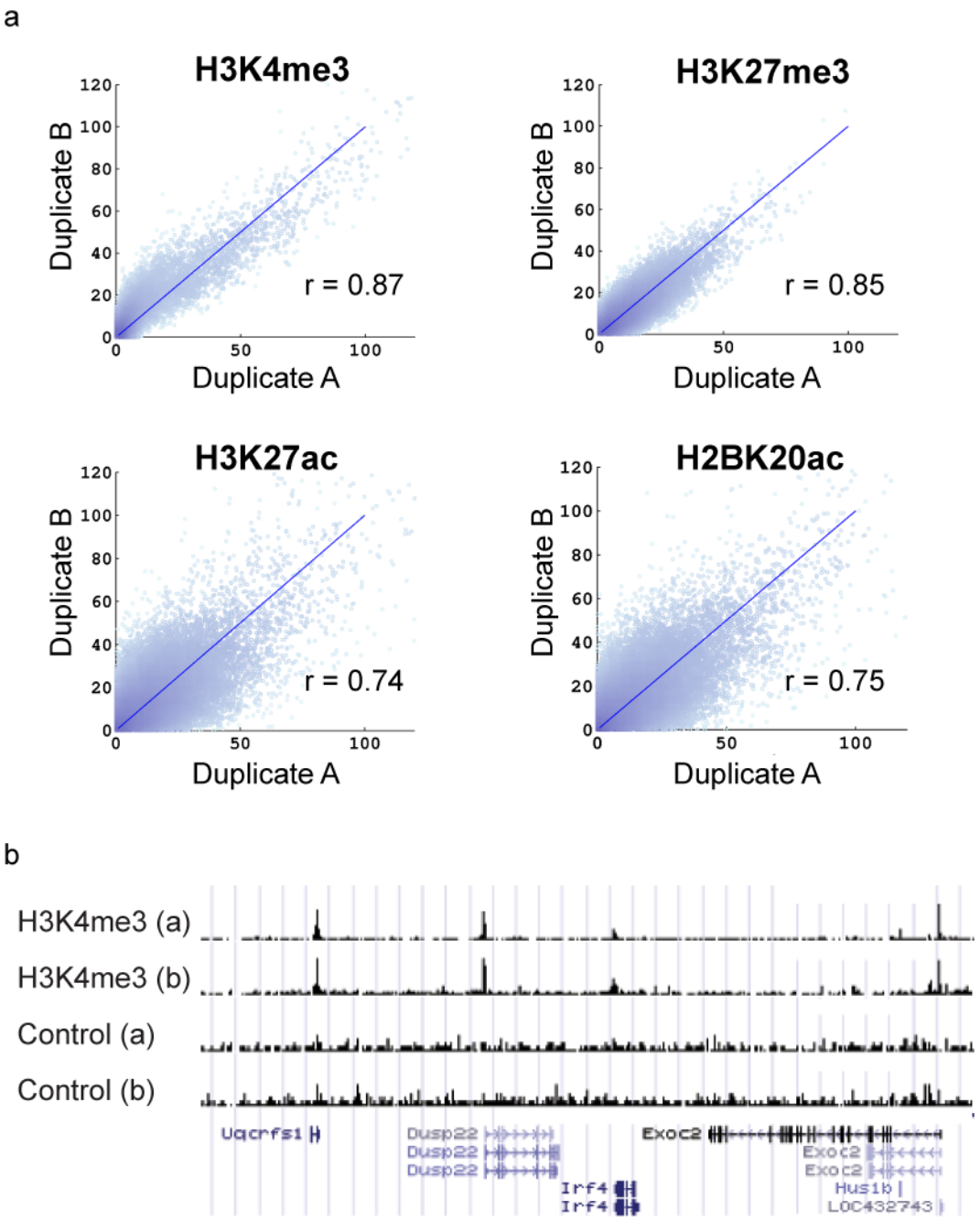

Figure 18: Reproducibility of small-scale ChIP-Seq. (a) Scatterplots of tag counts across 1-kb bins for ChIP duplicates of H3K4me3, H3K27me3, H3K27ac and H2BK20ac. (b) UCSC browser view illustrating ChIP-Seq signal profile for ChIP duplicates of H3K4me3 and controls.

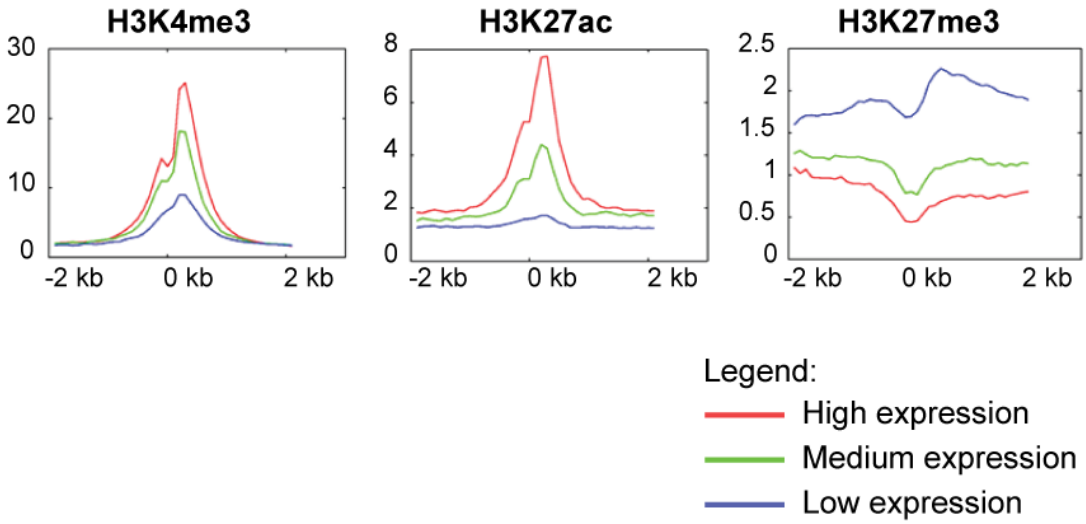

Figure 19: Relationship between histone mark occurrence and expression levels. Plotting the average ChIP-Seq profiles of H3K4me3, H3K27ac and H3K27me3 for genes with high (red), medium (green) and low (blue) expression. y-axis = fold relative to genome mean, x-axis = distance from transcription start site.

of H3K27me3, and the magnitude of difference among the different levels of expression was not as striking as H3K4me3 and H3K27ac. H3K27me3 may not be a good indicator of expression since bivalently marked regions are expressed at basal levels (Berstein et al., 2006). This could also signify that other repressive histone marks and perhaps DNA methylation may play more critical roles than H3K27me3 in suppressing gene expression. A dip was also observed at the TSS, a phenomenon which may be attributed to nucleosome depletion at transcription factor binding sites.

The ChIP-Seq peaks were viewed on the UCSC browser at specific genomic locations (Figure 20). Promoters are generally marked by H3K4me3 near the TSS. Pluripotency-associated genes such as *Tcl1*, *Zic3* and *Klf2* as well as germ cell-associated genes such as *Bbx* and *Dazl* exhibit active marks of H3K27ac and H2BK20ac but not H3K27me3. On the other hand, known bivalent genes such as *Dlx* and *Hox* genes are marked by both H3K4me3 and H3K27me3, but not H3K27ac and H2BK20ac. Interestingly, the *Esrrb* promoter region is already marked by bivalent markers although its transcript is expressed in E13.5 germ cells. However, it remains unclear if the *Esrrb* promoter is simultaneously marked by both H3K4me3 and H3K27me3 or this result represents a heterogenous cell population with the Esrrb promoter being marked by either histone modification.

Next, we wanted to focus on promoters that are specifically activated in germ cells. Germ cell-specific active promoters, marked by both H3K4me3 and H3K27ac, were identified by comparison with other cell types such as liver, ProB, neural progenitor cells (NPC) and mESCs (Figure 21a) (Creyghton et al., 2010). Genomic Regions Enrichment of Annotations Tool (GREAT) analysis was performed using whole

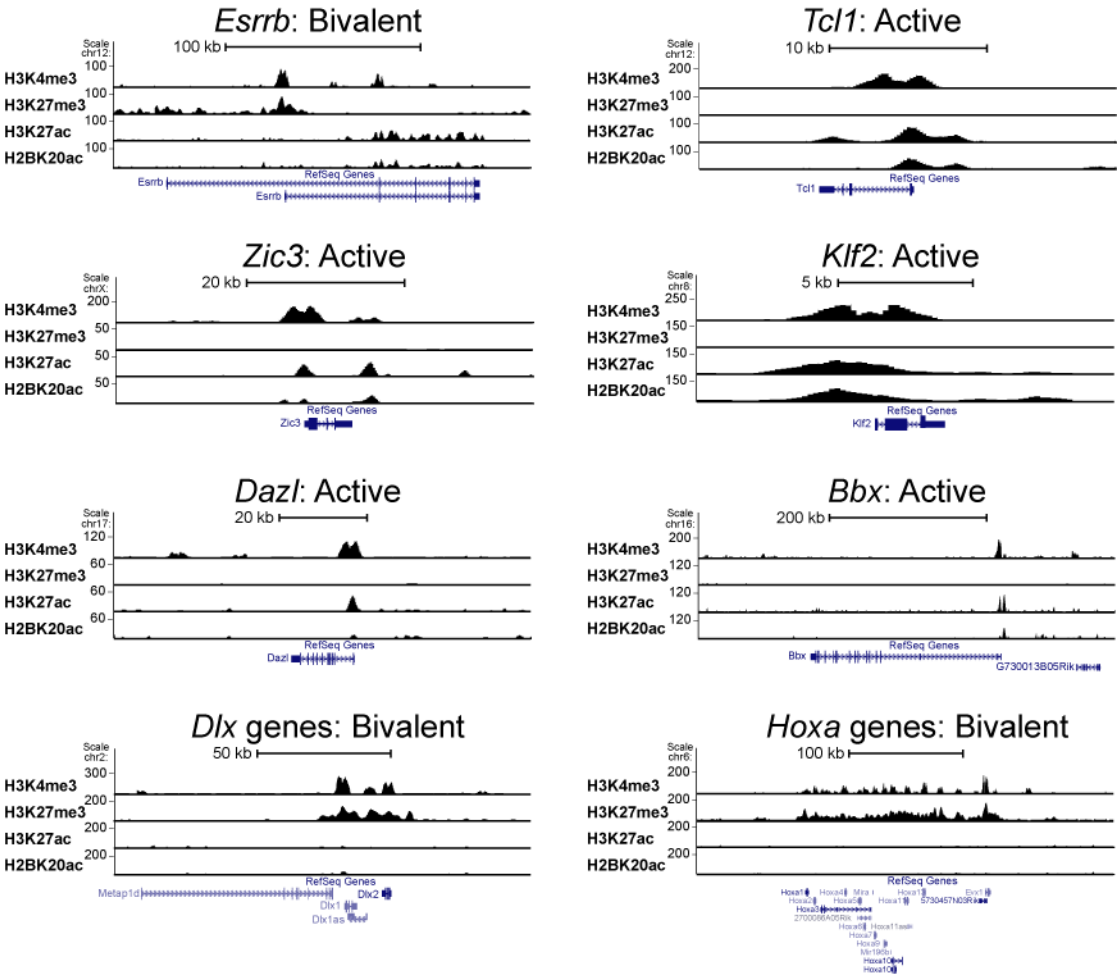

Figure 20: UCSC browser view of active and bivalent regions.

Active promoters are marked by H3K4me3 and H3K27ac (and H2BK20ac) whereas bivalent regions are marked by H3K4me3 and H3K27me3.

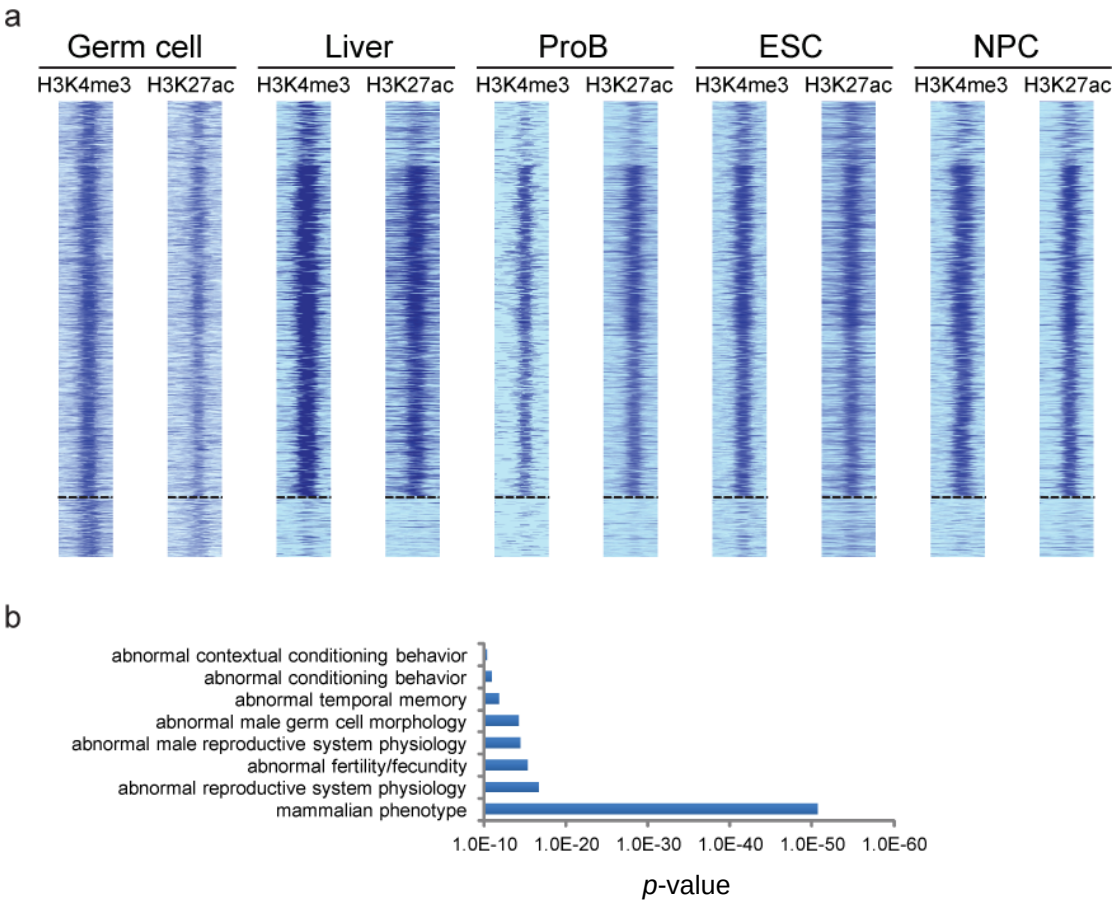

Figure 21: Discovery of E13.5 germ cell-secific active promoters with biological relevance. (a) Plot of H3K4me3 and H3K27ac ChIP-Seq signals, which are hallmarks indicative of active promoters. Germ cell-specific promoters are shown below the dotted line. (b) Enriched GO categories for mouse phenotype for genes located near E13.5 germ cell-specific active promoters (GREAT analysis: single nearest gene within 30 kb; FDR < 0.05; *q*-value < 9E-09).

mouse genome as background and taking positives as the nearest gene within 30 kb of the germ cell-specific promoters. By setting a false discovery rate (FDR) cut off as 0.05, genes located near germ cell-specific active promoters were found to be highly enriched (*q*-value < 9E-09) for mouse phenotypes related to reproduction and male fertility (Figure 21b).

Besides promoters, enhancers are also regulatory elements that are tightly regulated during development (Heintzman et al., 2009; Visel et al., 2009a). A total of 3494 active enhancers were identified based on H3K27ac and H2BK20ac peaks in both biological replicates and were not lying on any known gene promoter or H3K4me3 peak of any replicate. To validate these germ cell enhancers, 15 genomic fragments with lengths ranging from 500 bp to 1.7 kb were cloned downstream of the luciferase gene that is driven by the *Pou5f1* minimal promoter (Table 1; Figure 23a). These enhancer regions could be lying within a gene body, near a TSS or far from any known genes (Figure 22). Four random genomic regions were also included as controls. Luciferase assay was then performed in an embryonic germ cell line, Tg2-EGC (Labosky et al., 1994) and 293T to check for enhancer activities. Eleven out of 15 tested regions (73.3%), but none of the 4 control regions, had more than 2-fold luciferase activity relative to the minimal promoter-only control ($p < 0.05$) (Figure 23b-c; Table 2-3). The enhancer property of these 11 regions was also specific to Tg2-EGC since luciferase activity was not detected in transfected 293T control cell type.

Given that germ cells resemble ESCs in several aspects such as expression of many pluripotency regulators, and early germ cells can be converted to pluripotent EGCs by

Table 1: Genomic locations of the tested enhancer regions.

Construct	Genomic location (Mm8 reference genome)	Length (bp)
Control1	chr7:43,750,426-43,751,690	1265
Control2	chr16:23,458,410-23,459,450	1041
Control3	chr5:123,077,975-123,078,987	1013
Control4	chr11:33,431,734-33,432,779	1046
Region1	chr1:121,313,475-121,314,580	1106
Region2	chr3:101,539,183-101,540,367	1185
Region3	chr3:34,956,203-34,956,977	775
Region4	chr17:83,659,727-83,660,321	595
Region5	chr11:77,732,994-77,734,498	1505
Region6	chr3:101,568,225-101,569,063	839
Region7	chr15:76,884,566-76,885,957	1392
Region8	chr14:63,361,450-63,362,945	1496
Region9	chr4:131,920,924-131,922,490	1567
Region10	chr5:131,774,542-131,775,774	1233
Region11	chr7:134,517,582-134,519,272	1691
Region12	chr4:118,523,908-118,525,092	1185
Region13	chr1:134,678,182-134,679,648	1467
Region14	chr16:17,133,278-17,134,396	1119
Region15	chr10:44080072-44080871	800

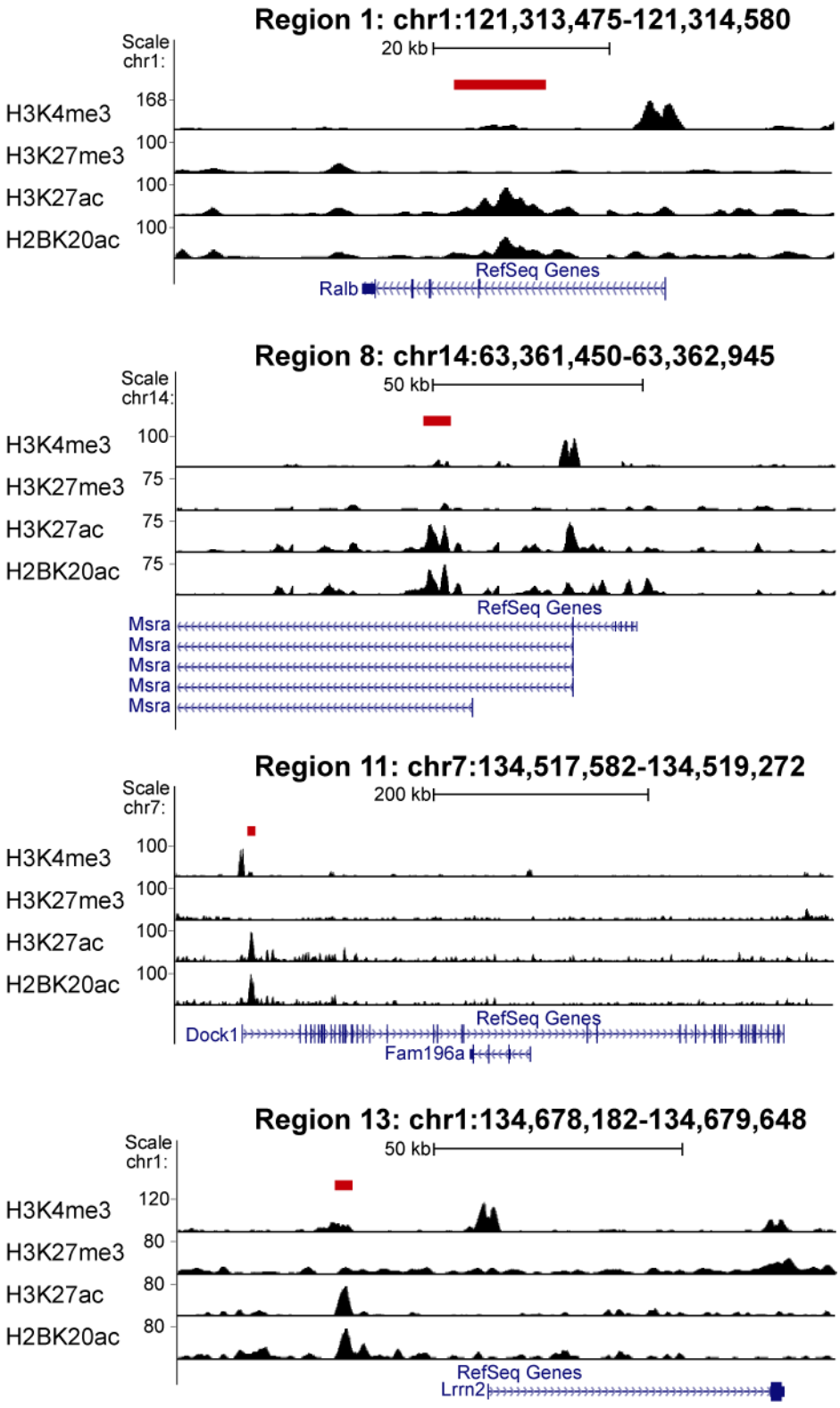

Figure 22: UCSC browser view of tested enhancer regions. The red horizontal bar indicates position of the enhancer region.

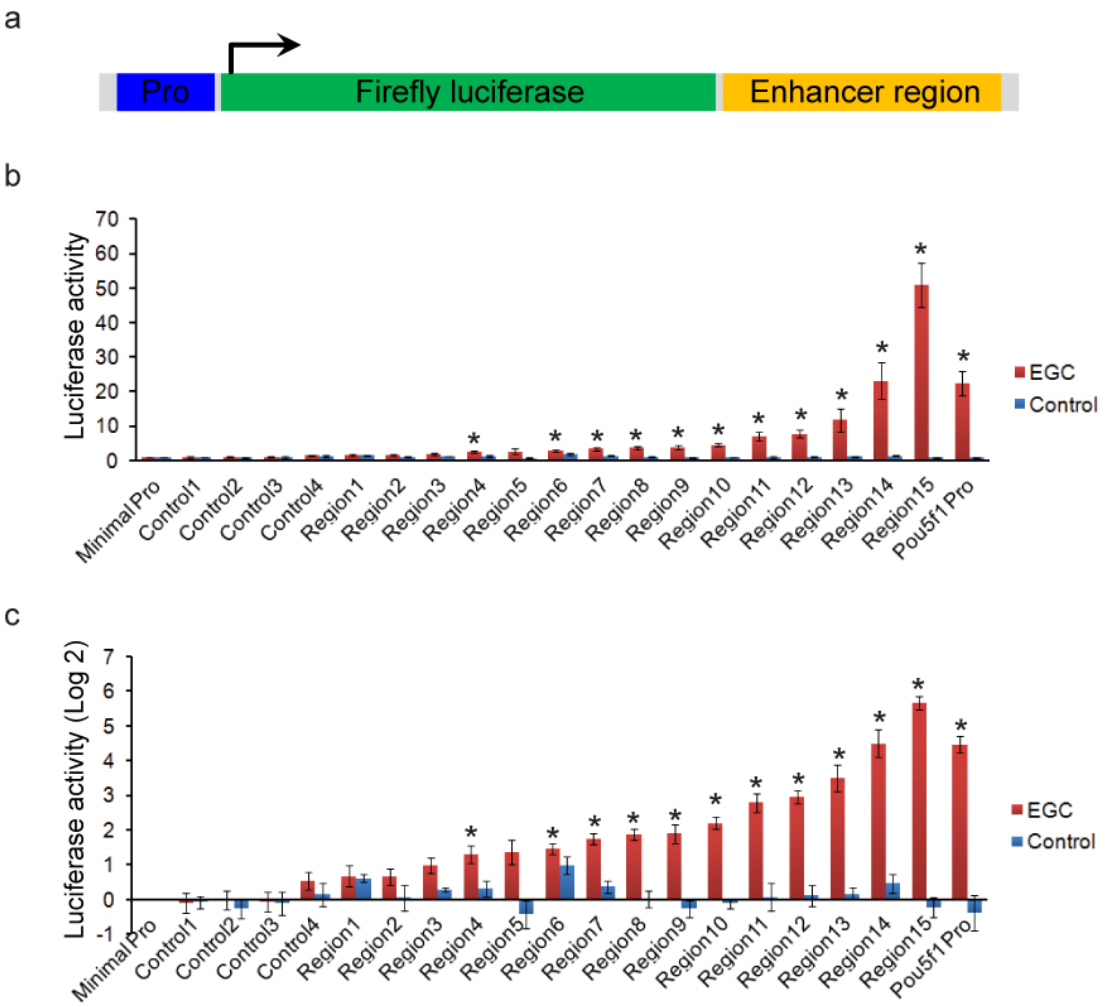

Figure 23: Luciferase assay validation of enhancers in EGC and 293T control. (a) Schematic of modified pGL3 vector that contains a 500-bp *Pou5f1* minimal promoter (Pro). The tested enhancer regions were cloned downstream of the firefly luciferase reporter gene. (b) Luciferase activity is expressed as a ratio relative to vector containing the minimal promoter only (Minimal Pro). The 2.3-kb *Pou5f1* promoter (*Pou5f1* Pro) which includes both distal and proximal enhancer serves as a positive control. Data represents the mean ± s.e.m. of five independent experiments (n = 5). Asterisks mark regions with at least 2-fold relative luciferase activity (Student's T-test, One-tailed p-value < 0.05). (c) Log2-transformed version of (b).

Table 2: Luciferase results of tested enhancer regions in EGCs.

Fold enrichment	Average (n = 5)	Standard Deviation	P-value (One-tail)
Minimal Pro	1.00	0.00	N.A.
Control1	0.94	0.18	1.06E-04
Control2	1.01	0.18	1.20E-04
Control3	0.98	0.18	1.02E-04
Control4	1.47	0.23	3.27E-03
Region1	1.63	0.34	3.58E-02
Region2	1.59	0.25	1.04E-02
Region3	2.00	0.31	4.92E-01
Region4 *	2.52	0.45	3.03E-02
Region5	2.65	0.76	6.37E-02
Region6 *	2.79	0.30	2.18E-03
Region7 *	3.35	0.34	4.26E-04
Region8 *	3.70	0.37	2.64E-04
Region9 *	3.79	0.71	2.43E-03
Region10 *	4.61	0.56	2.31E-04
Region11 *	6.98	1.35	5.80E-04
Region12 *	7.78	1.02	1.12E-04
Region13 *	11.67	3.32	1.43E-03
Region14 *	23.07	5.30	4.41E-04
Region15 *	50.94	6.44	3.52E-05
Pou5f1 Pro	22.29	3.56	1.09E-04

Table 3: Luciferase results of tested enhancer regions in 293T control.

Fold enrichment	Average (n = 5)	Standard Deviation	P-value (One-tail)
Minimal Pro	1.00	0.00	N.A.
Control1	0.96	0.12	1.87E-05
Control2	0.84	0.17	5.26E-05
Control3	0.95	0.22	2.37E-04
Control4	1.14	0.29	1.32E-03
Region1	1.53	0.13	5.70E-04
Region2	1.06	0.25	5.41E-04
Region3	1.22	0.06	5.31E-06
Region4	1.26	0.19	5.03E-04
Region5	0.76	0.20	8.02E-05
Region6	2.00	0.37	4.90E-01
Region7	1.30	0.16	2.80E-04
Region8	1.02	0.16	7.29E-05
Region9	0.84	0.14	2.53E-05
Region10	0.93	0.09	6.92E-06
Region11	1.09	0.31	1.37E-03
Region12	1.10	0.23	4.82E-04
Region13	1.12	0.13	6.10E-05
Region14	1.40	0.26	3.19E-03
Region15	0.87	0.18	7.01E-05
Pou5f1 Pro	0.80	0.23	1.55E-04

altering culture conditions alone, one attractive hypothesis is that germ cells may also be epigenetically more similar to ESCs. Out of 3494 active enhancers in E13.5 male germ cells, 19.6% overlap with active enhancers in mESC (Figure 24). The overlap of active enhancers with ESC is 11.6%, 12.4% and 19% for liver, ProB and NPC respectively. By filtering active enhancers in liver, ProB, ESC and NPC, germ cell-specific enhancers were discovered (Figure 25a). GREAT analysis was performed using whole mouse genome as background and positives as the nearest gene within 30 kb of the germ cell-specific enhancers. Taking FDR < 0.05 and *q*-value < 0.001, genes neighboring germ cell-specific enhancers are enriched for mouse phenotypes such as gametogenesis, reproduction and testicular abnormalities (Figure 25b), suggesting biological relevance of the identified germ cell enhancers.

In summary, our modified small-scale ChIP protocol enabled the genome-wide mapping of histone modifications from small numbers of purified germ cells, from which invaluable information could be retrieved. The combinatorial occurrences of histone modifications reveal the nature (promoter vs enhancer) and status (active vs inactive/ poised) of regulatory elements, and can be useful for predicting gene expression. In addition, by comparison with publicly available ChIP-Seq data for other cell types (Creyghton et al., 2010), germ cell-specific enhancers with implicated biological relevance were identified. Given that enhancers serve as important regulatory sites for transcription factors, further analyses may help reveal key germ cell regulators.

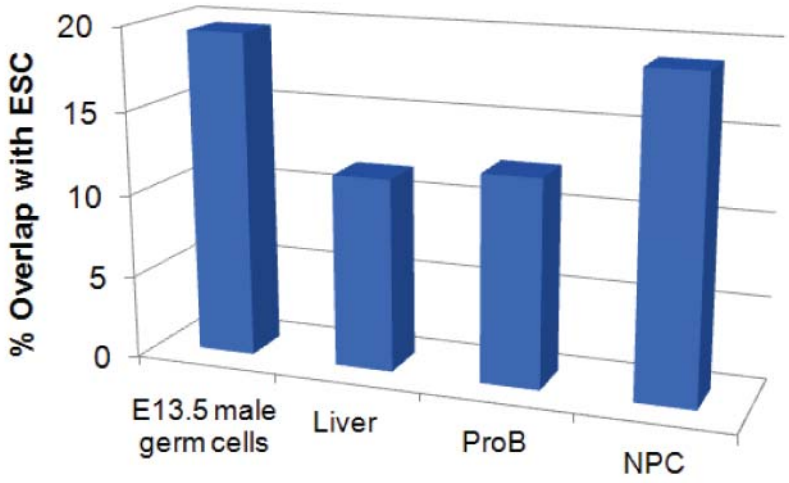

Figure 24: Overlap of cell type-specific enhancers with ESC enhancers.

The overlap of E13.5 germ cell-specific enhancers, Liver-specific enhancers, ProB-specific enhancers and NPC-specific enhancers with ESC enhancers is expressed as a percentage of the respective cell type-specific enhancers. Data from other somatic cell types, besides E13.5 germ cells, is derived from previously published data (Creyghton et al., 2010).

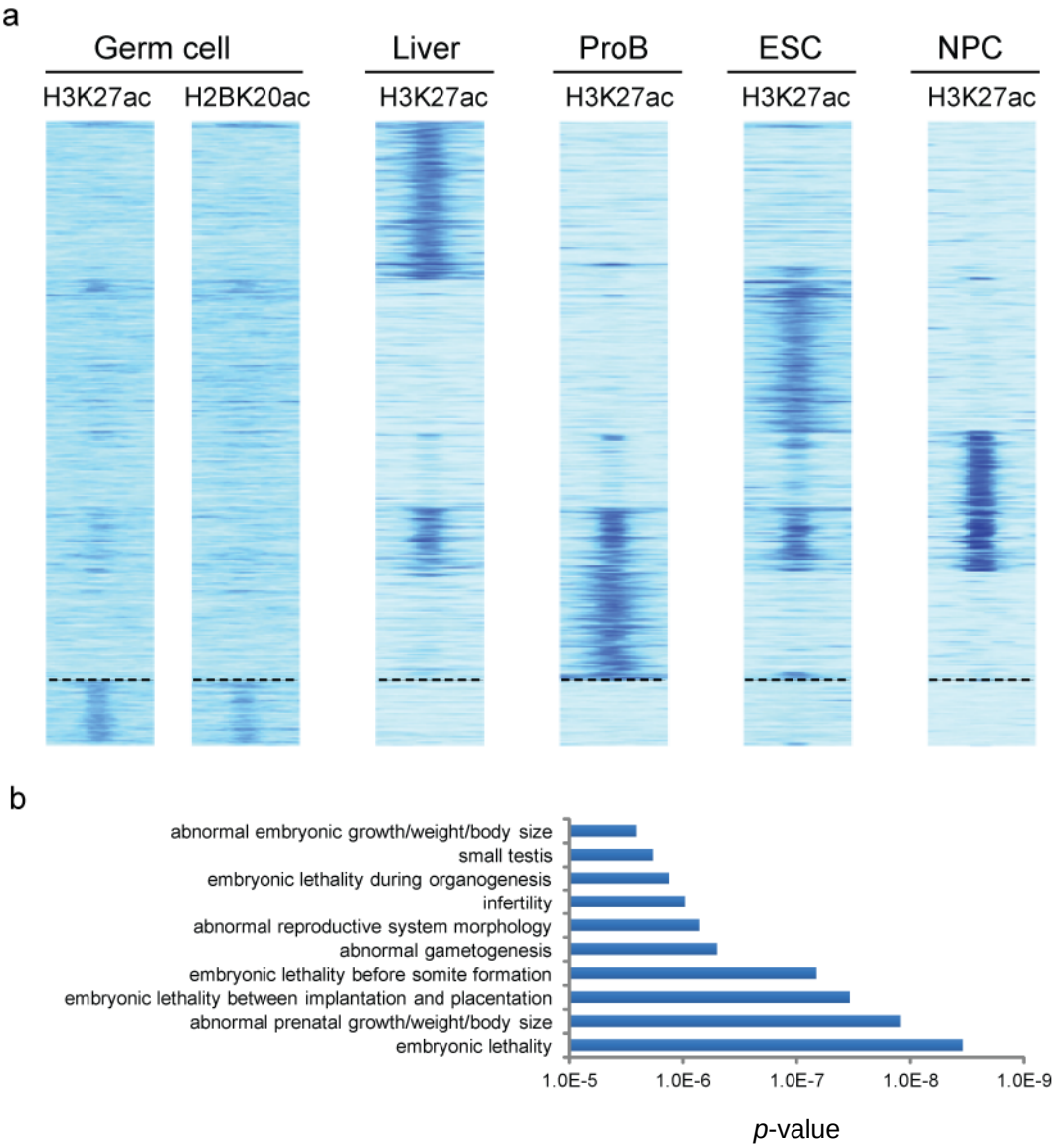

Figure 25: Discovery of E13.5 germ cell-specific enhancers with biological relevance (a) Plot of H3K27ac and H2BK20ac ChIP-Seq signals, which are hallmarks indicative of active enhancers. Germ cell-specific enhancers are shown below the dotted line. (b) Enriched GO categories for mouse phenotype for genes located near E13.5 germ cell-specific enhancers. (GREAT analysis: single nearest gene within 30 kb; FDR < 0.05; *q*-value < 0.001).

9.4. Discovery of germ cell regulators

9.4.1. Potential regulators of germ cell differentiation and "pluripotential"

To discover enriched motifs found at germ cell enhancers, the Multiple EM for Motif Elicitation (MEME) program was applied. Our *de novo* motif search led to the discovery of a 15-bp palindromic motif (Figure 26) that does not match any known transcription factor motif. It is possible that the transcription factor that binds this motif at germ cell enhancers is yet to be discovered. Alternatively, such as transcription factor may have different binding preference in germ cells compared to the motif that is published in the MEME database. We then attempted to search for potential germ cell regulators by finding comparative enrichment of known transcription factor motifs in the cell type-specific enhancers (Table 4). Notably, the top enriched motifs that emerged from our analysis have known functions for the respective cell types. For example, HNF and CEBPA for liver enhancers (Cereghini, 1996), ETS family motifs for ProB enhancers (Maroulakou and Bowe, 2000) and AP2 for NPC enhancers (Schmidt et al., 2011). Similarly, E2F, Klf4, Esrrb and Nr5a2 motifs were among the top motifs enriched at ESC enhancers. Motifs that were more enriched at E13.5 male germ cell enhancers include FoxP1 and Dmrt. Interestingly, FoxP1 is a member of the forkhead box (FOX) family of transcription factors which play important roles in regulating cell type-specific transcriptional programs of diverse cell types (Hannenhalli and Kaestner, 2009). An ESC-specific isoform of FoxP1 was recently shown to participate in iPSC reprogramming by suppressing the expression of differentiation genes and stimulating pluripotency-associated genes

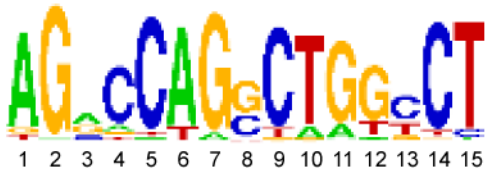

Figure 26: *De novo* motif discovery using germ cell enhancers.

The position weight matrix (PWM) of the enriched motif found at germ cell enhancers.

Table 4: Top enriched motifs at cell type-specific enhancers.

Germ cells	Liver	ProB	ESC	NPC
FOXP1	PLAG1	ETF	SP1	HIC1
FAC1	HNF1	Ets-1	E2F	NF-1
DMRT3	HNF4	c-Ets-1	KLF4	AP-2
RREB1	CEBPA	RSRFC4	Esrrb	AP-4
FOX	NR2F1	ZF5	p53	Adf-1
Pbx-1	Foxa2	PU.1	Nr5a2	TLX1:NFIC
Tra-1	NFIC	ISCBP	MAF	ZF5
Cdx	GATA3	GABPA	RAV-1	LBP-1
GBF	Freac-4	Nrf-1	Oct4	Zfx

such as *Pou5f1*, *Nanog*, *Nr5a2* and *Gdf3* (Gabut et al., 2011). Members of the doublesex and mab-3-related transcription factor (Dmrt) family are known for their conserved functions in sex determination (Hong et al., 2007). Strikingly, Dmrt motifs were also among the top hits when we searched for transcription factors whose motif is anti-correlated with H3K27me3 at ESC-bivalent sites (Figure 27). This result suggests that Dmrt family members may be important in regulating genes that are poised in ESCs but become activated in germ cells.

To find transcription factors that are crucial in regulating transcriptional changes and potentially play a role in mediating developmental transition of germ cells during the 11.5 to 15.5 d.p.c period, motif enrichment analysis was also applied to regulatory elements located near differentially expressed genes. We first clustered genes based on microarray expression, and found a set of genes whose expression is downregulated in E15.5 germ cells as compared to E11.5 germ cells (Class 1), and another cluster of genes whose expression is upregulated in E15.5 germ cells in comparison to E11.5 germ cells (Class 2) (Figure 28a). For both classes of genes, we selected H3K27ac and H2BK20ac peaks which were located within 10 kb of their transcriptional start site. Upon motif enrichment analysis, we found that peaks located around Class 2 genes had higher occurrence rate for transcription factors such as Egr1, Egr4, Znf219 and MovoB (Figure 28c), some of which are known to be important for male germ cell differentiation (Hogarth et al., 2010; Tourtellotte et al., 1999; Tourtellotte et al., 2000). *Egr1*, *Egr4* and *Znf219* are upregulated from E11.5 to E15.5, although the upregulation of *Znf219* is modest (Figure 29b); *MovoB*, also known as *Ovol2*, is expressed at slightly higher levels at E13.5 but is quickly downregulated by E15.5 (Figure 29b). Interestingly, Znf219 is also a protein

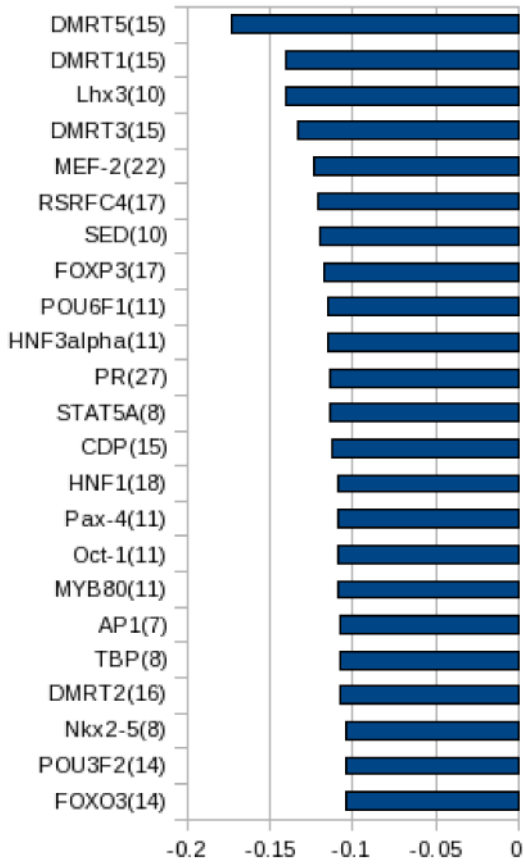

Figure 27: Motifs found to be anti-correlated to H3K27me3.

Using ESC-bivalent regions (i.e. regions marked by both H3K4me3 and H3K27me3 in ESC), the correlation of motif enrichment to occurrence of H3K27me3 in E13.5 germ cells was computed. The negative correlation (anti-correlation), as indicated by the x-axis of the plot, of Dmrt motifs to H3K27me3 in E13.5 germ cells suggests that Dmrt1-bound ESC-bivalent regions lose their H3K27me3 mark.

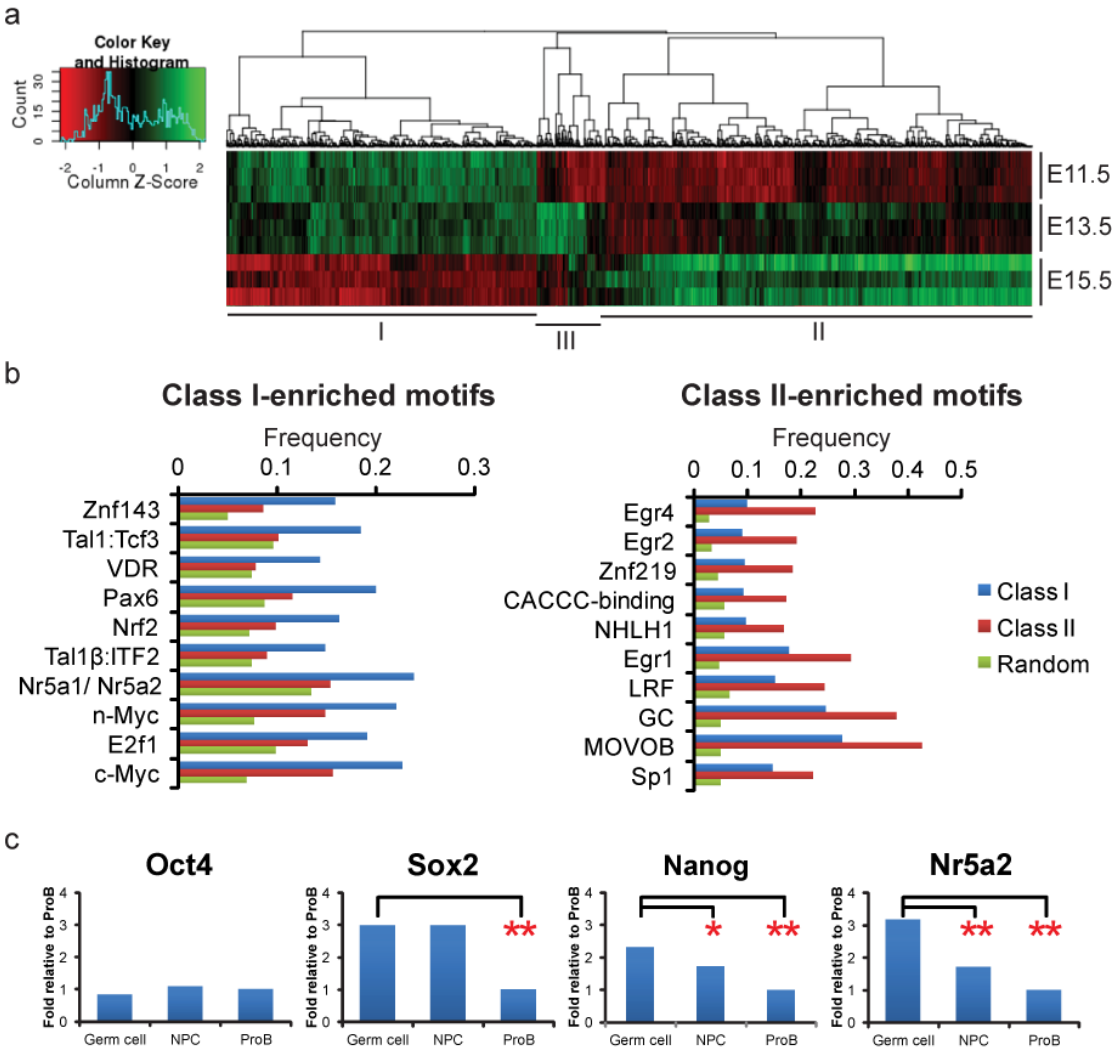

Figure 28: Motif analysis of differentially expressed genes. (a) Microarray heatmaps of Class 1 and Class 2 genes. (b) Enriched motifs for Class 1 and Class 2 genes. (c) Transcription factor binding at cell type-specific enhancers. The frequencies of transcription factor binding at cell type-specific enhancers are expressed as fold relative to ProB's frequency. Asterisk indicates significant (* $p < 0.05$; ** $p < 0.005$) enrichment of binding at germ cell-specific enhancers compared to Liver, NPC and ProB-specific enhancers.

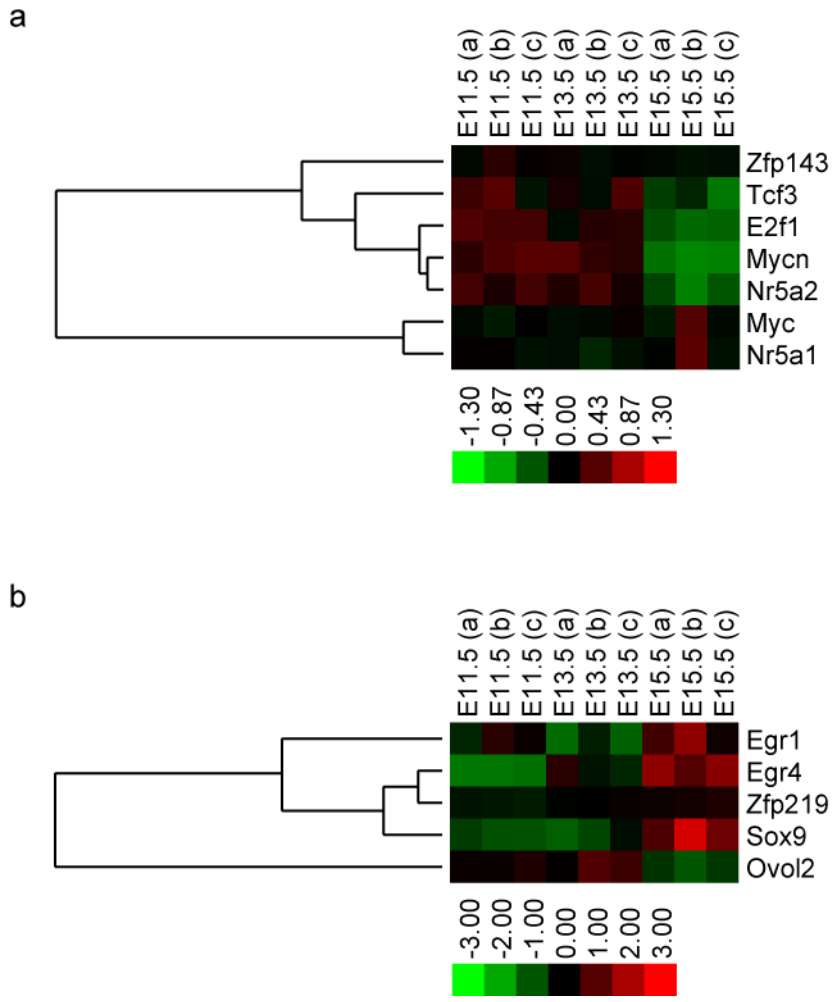

Figure 29: Microarray heatmaps of transcription factors with motifs enriched at (a) Class1 and (b) Class2 genes. Biological triplicate data (a,b and c) are shown for each timepoint.

interaction partner of Oct4 in mESCs (Pardo et al., 2010; van den Berg et al., 2010; Wang et al., 2006), and associates with Sox9 to regulate chondrocyte differentiation (Takigawa et al., 2010). Incidentally, Sox9 is a central male sex-determining factor (da Silva et al., 1996; Kocer et al., 2009), which is highly expressed in non-germ cells and upregulated in E15.5 germ cells (Figure 29b). As such, it would be interesting to further examine whether Znf219 also interacts with Sox9 in the gonad, and if Znf219 is important for germ cell differentiation. Importantly, the peaks located near Class 1 genes had higher frequency of motifs for transcription factors that have been functionally associated to ESC maintenance, iPSC reprogramming or cell proliferation (Figure 28b), as exemplified by Znf143 (Chen et al., 2008a), Tcf3 (Tam et al., 2008; Wray et al., 2011; Yi et al., 2011; Yi et al., 2008), Nr5a2/Nr5a1 (Guo and Smith, 2010; Heng et al., 2010), Myc (Cartwright et al., 2005; Lin et al., 2009; Singh and Dalton, 2009; Smith et al., 2010) and E2F1 (Chen et al., 2008b; Chong et al., 2009). Microarray data also reveals that these factors themselves are downregulated by E15.5 (Figure 29a). Therefore, shutting down the pluripotency-related network seems to be a major theme for germ cells during the E11.5 to E15.5 d.p.c. period.

The similarity between the motifs of Nr5a1 and Nr5a2 suggests that either one of them could be binding the H3K27ac/ H2BK20ac-marked sites to maintain "pluripotential" of early germ cells. To corroborate the above analysis, we compared the overlap of enhancers from different cell types with transcription factor ChIP-Seq peaks from mESC (Chen et al., 2008b). Sox2 motif is enriched at both E13.5 germ cell and NPC enhancers (Figure 28c). Enrichment of Sox2 binding at NPC enhancers may not be surprising given that both Sox2 has been implicated in neural differentiation (Ferri et al., 2004). Interestingly, Nanog and Nr5a2 ChIP-Seq peaks

Table 5: Transcription factor binding (fold) at cell type-specific enhancers.

The frequencies of transcription factor ChIP-Seq peaks at cell type-specific enhancers of E13.5 male germ cells (MGC), NPC and ProB are expressed as a fold relative to the latter's frequency.

Transcription factor	Fold relative to ProB frequency		
	MGC	NPC	ProB
Oct4	0.83	1.08	1.00
Sox2	2.99	2.99	1.00
Nanog	2.32	1.73	1.00
Nr5a2	3.18	1.72	1.00

Table 6: Two proportion p-values and test statistics for transcription factor binding at cell type-specific enhancers.

The p-value and Z-score statistics to test for significant difference between the transcription factor binding frequencies at E13.5 male germ cell (MGC) compared to either NPC or ProB. Significant differences (p-value < 0.05) are indicated by red font.

Transcription factor	MGC vs NPC		MGC vs ProB	
	p-value	(Z-score)	p-value	(Z-score)
Oct4	8.64E-02	(-1.71)	1.71E-01	(-1.37)
Sox2	9.85E-01	(0.02)	9.42E-13	(7.14)
Nanog	1.20E-02	(2.51)	5.70E-14	(7.51)
Nr5a2	1.70E-03	(3.14)	4.32E-10	(6.24)

were more enriched at E13.5 germ cell enhancers in comparison to enhancers of NPC and ProB (p-value < 0.05) (Figure 28c; Table 5-6). Nanog is expressed in migratory and mitotic post-migratory germ cells (Yamaguchi et al., 2005) and has been reported to be important for developmental progression of male germ cells beyond E11.5 (Chambers et al., 2007). Taken together, this result suggests that Nr5a2 may potentially be important for germ cell regulation.

9.4.2. Dmrt1

Members of the Dmrt family of transcription factors persistently emerge in our search for germ cell regulators. Highly related to the *doublesex* gene of Drosophila and the *mab-3* gene of Caenorhabditis elegans, Dmrt family members are characterized by the conserved DM (*dsx* and *mab-3*) DNA binding domains and roles in sexual differentiation (Hong et al., 2007). Among the eight Dmrt members in mammals, Dmrt1 is the member that is best known for its gonadal-specific expression and functions in both sex determination as well as male germ cell differentiation in a myriad of species including mouse, fish, birds and amphibians (Hong et al., 2007; Raymond et al., 2000). In humans, deletion of a region on chromosome 9 where Dmrt1, Dmrt2 and Dmrt3 are located causes 46, XY gonadal dysgenesis and sex reversal (Raymond et al., 1999). Other than Dmrt1, several other family members such as Dmrt3, Dmrt4, Dmrt7 and Dmrt8 were previously reported to be expressed in the embryonic gonad (Kim et al., 2003; Veith et al., 2006). However, it is unclear if all of these factors are expressed in the germ cells or the gonadal somatic cells such as Sertoli or Leydig cells.

Microarray data, validated by qPCR, reveals that *Dmrt1* and *Dmrtc1c* (also known as *Dmrt8*) are the only two family members that are more highly expressed in E13.5 germ cells (Figure 30a,b). Curiously, their transient expression at E13.5 decreased to non-germ cell levels by E15.5. Another family member, *Dmrt7*, is gradually upregulated and is expressed significantly more than non-germ cells on 15.5 d.p.c. *Dmrt7* is required for male but not female gametogenesis given that homozygous knockout males are infertile in contrast to fertile females (Kim et al., 2007b). However, *Dmrt7* is only required at a later stage of gametogensis as *Dmrt7* mutant germ cells arrest postnatally at the pachytene stage of meiosis (Kim et al., 2007b). On the other hand, Dmrt8 is unique in that it does not have the DM domain that the other seven family members possess (Veith et al., 2006), and was proposed to dimerize with other Dmrt family members to act as either co-regulator or dominant-negative regulator (Veith et al., 2006). Taking both germ cell expression and literature into consideration, it is likely that Dmrt1 is the family member that binds to active enhancers in E13.5 male germ cells.

Previous studies have shown that *Dmrt1* expression is dynamically regulated in male and female gonads in a sex-dependent manner. In the embryo, *Dmrt1* is expressed in both male and female germ cells with a peak in expression level at E13.5 followed by downregulation at E15.5 (Lei et al., 2007). However, *Dmrt1* is increasingly expressed in male but not female gonadal somatic cells from E12.5 onwards, and such sexually dimorphic expression is maintained after birth when *Dmrt1* is expressed in the testis but not the ovary (Lei et al., 2007). Interestingly, *Dmrt1* is upregulated in germ cells again on 0.5 days post partum (d.p.p) in the testis only (Lei et al., 2007). Given that the expression pattern of *Dmrt1* in germ cells corresponds to the proliferative stages

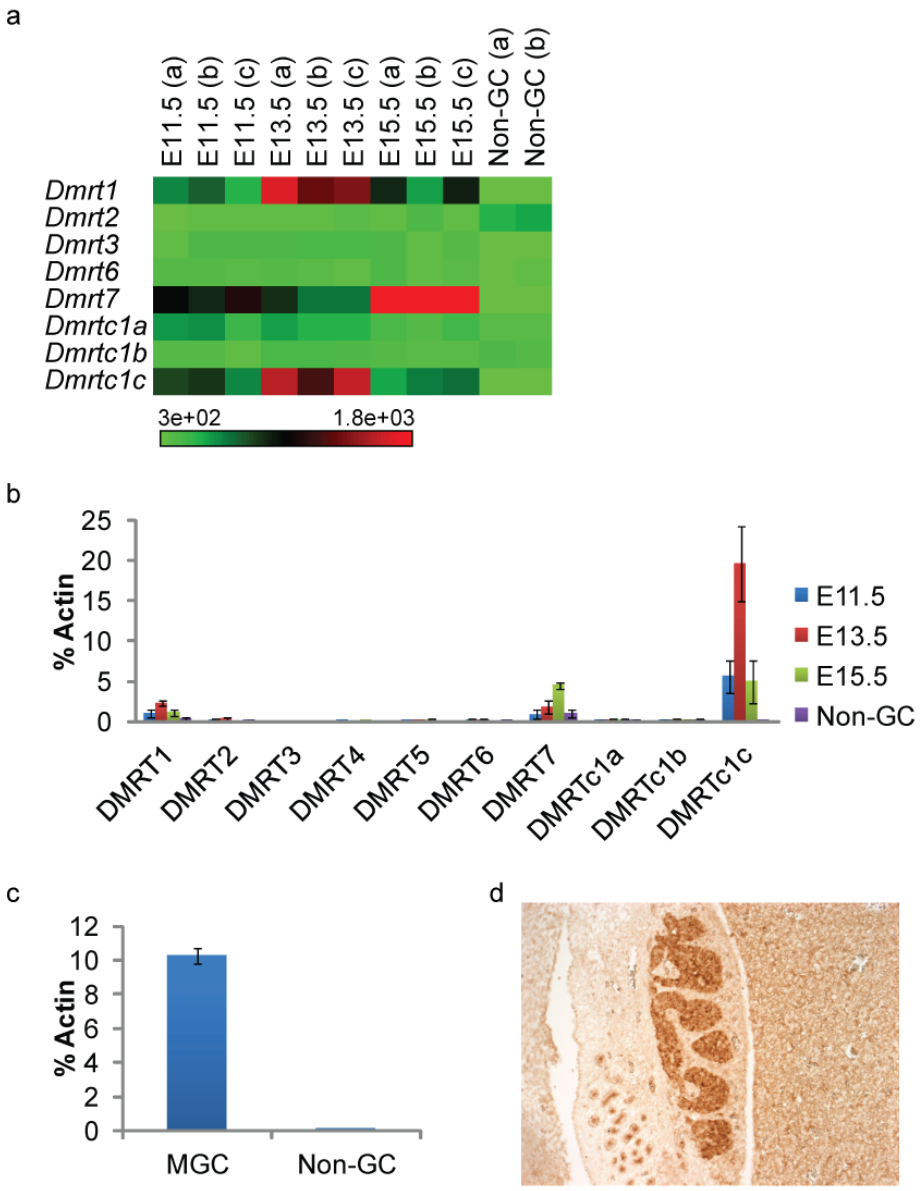

Figure 30: Expression of Dmrt1 and Dmrt family members in germ cells. (a) Microarray heatmaps and (b) qPCR results for Dmrt family members in male germ cells (MGC: E13.5 male germ cells) and non-germ cells (Non-GC). Data in (b) represents the mean ± s.e.m. of three biological replicates (n = 3). (c) qPCR results for Dmrt1 using non-amplified cDNA of E13.5 male germ cells (MGC) and non-germ cells . Data represents the mean ± s.e.m. of two biological replicates (n = 2). (c) IHC staining of E13.5 male gonadal sections using anti-Dmrt1 antibody.

of germ cells after arrival at the gonad, both before and after birth, it seems likely that Dmrt1 may regulate cell cycle processes. In agreement with this postulation, loss of *Dmrt1* results in high incidence of teratoma formation due to ectopic expression of pluripotency gene *Sox2* as well as failure of germ cells to enter mitotic arrest (Krentz et al., 2009). Notably, this phenotype was only observed in the 129 mouse strain, but not other mouse strains. In humans, Dmrt1 has also been associated with testicular germ cell tumor susceptibility (Kanetsky et al., 2011; Turnbull et al., 2010). Other than a role in mediating cell cycle arrest, Dmrt1 is important in directing appropriate sex differentiation. In female germ cells, Dmrt1 promotes oogenesis by activating *Stra8* and inducing meiosis (Krentz et al., 2011). In the postnatal spermatogonia, Dmrt1 plays an opposite role whereby it represses *Stra8* to suppress meiosis and activates the spermatogonial differentiation factor *Sohlh1* (Matson et al., 2010). In sertoli cells of the testis, Dmrt1 antagonizes Foxl2 in order to maintain a sertoli instead of granulosa cell identity (Matson et al., 2010). Collectively, these studies suggest that Dmrt1 plays pivotal context-dependent roles in both germ cells and gonadal somatic cells, and regulates mitosis/ meiosis as well as other aspects sex differentiation differentially in male and female germ cells (Kim et al., 2007a).

To confirm that *Dmrt1* is expressed in the germ cells purified from the *Pou5f1-GFP* (Strain BL6 x CD1) transgenic mice, realtime qPCR was performed using non-amplified E13.5 cDNA and indeed *Dmrt1* was much enriched (260-fold) in E13.5 male germ cells compared to non-germ cells (Figure 30c). Dmrt1 protein expression in E13.5 male germ cells was also verified by immunohistostaining (IHC) using rabbit serum that was raised against a Dmrt1 antigen (Figure 30d). Given its repression on pluripotency genes and inhibition of cell cycle (Krentz et al., 2009), it is surprising to

find that Dmrt1 is expressed in mESC, as its protein could be detected from both whole cell lysate (WCL) and isolated chromatin of E14 mESC by Western blot (Figure 31a). To investigate any potential roles of Dmrt1 in ESC, shRNA oligos were designed and cloned into the pLKO.1 vector. Transfected E14 mESCs exhibited *Dmrt1* knockdown of more than 75% (Figure 31b,c). After 4 days of *Dmrt1* knockdown, either no change or a slight upregulation was observed for pluripotency regulators such as *Pou5f1*, *Sox2*, *Nanog* and *Nr5a2*. *Dmrt1* knockdown cells also appeared morphologically similar to undifferentiated mESC. Therefore, Dmrt1 does not appear to play a role in ESC maintenance.

The transient upregulation of *Dmrt1* in male germ cells at E13.5 d.p.c. (Figure 30a,b) suggests that Dmrt1 may play a role in mediating the transition towards male germ cell differentiation. In an attempt to drive germ cell differentiation, *Dmrt1* was overexpressed under the control of the constitutive cytomegalovirus (CMV) enhancer/chicken β-actin (CAG) promoter in E14 mESC. Within 3 days of transfection, *Dmrt1*-overexpressing mESC displayed increased cell death and the remaining attached cells appeared differentiated (Figure 32a). This phenotype is accompanied by downregulation of pluripotency regulators *Pou5f1*, *Sox2* and *Nanog* (Figure 32b). Germ cell genes such as *Dazl*, *Ddx4* and *Stra8* were initially upregulated up to 200% of control levels after 1 day of overexpression (Figure 32c). However, this upregulation was soon diminished, and expression of *Dazl* and *Stra8* were reduced by Day 4. Probing further, I found that markers of other lineages especially trophectoderm (*Cdx2* and *Hand1*) and endoderm (*Gata4* and *Gata6*) were conspicuously upregulated (Figure 32d). The upregulation of other lineage markers may be an indirect consequence of Dmrt1's repression of pluripotency factors. This

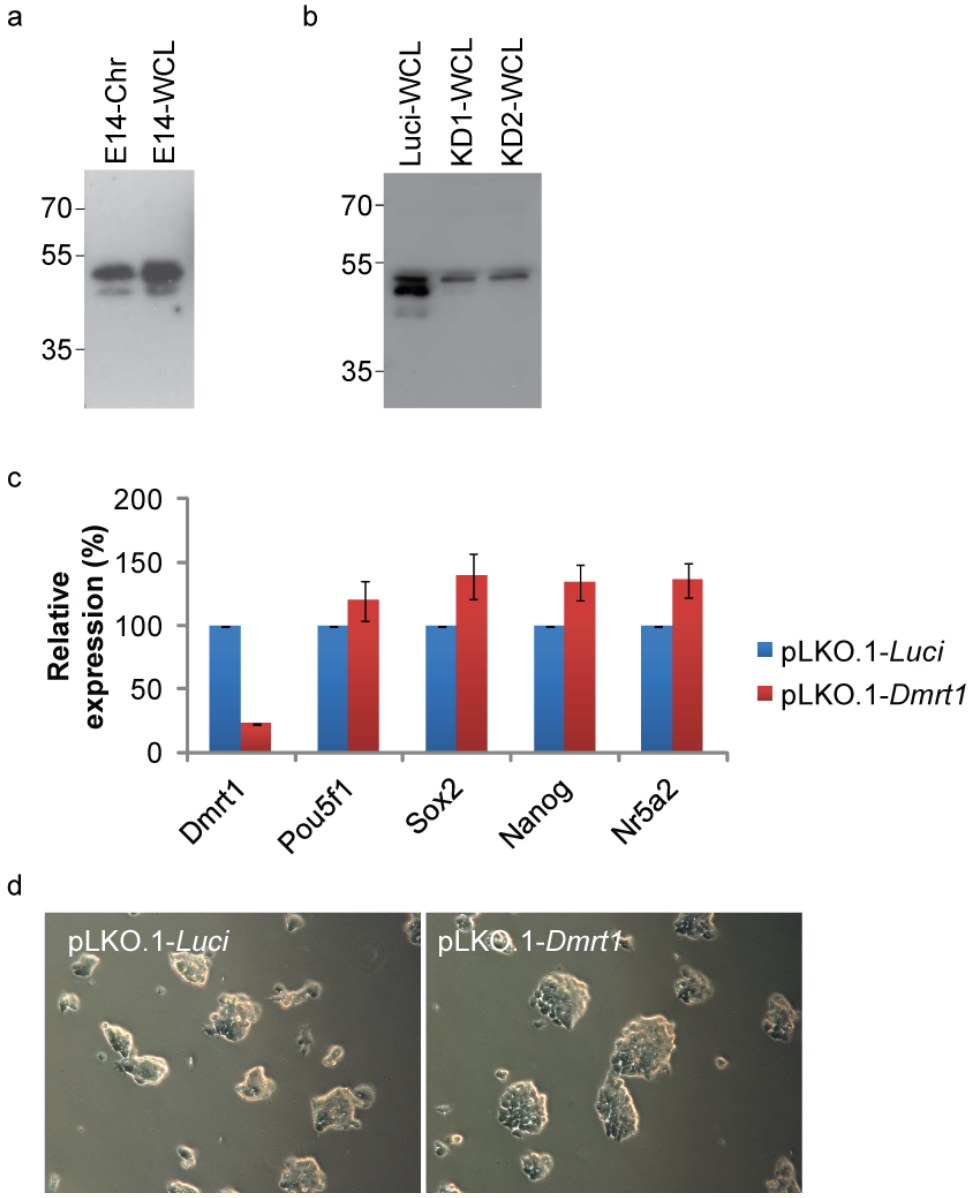

Figure 31: Dmrt1 is expressed but not required in mESC. (a) Western blot against Dmrt1 in E14 mESC chromatin (Chr) or whole cell lysate (WCL). (b) Western blot, using anti-Dmrt1 serum, was performed on transfected E14 mESC WCL after 3 days of selection. Cells were transfected with two shRNAs against *Dmrt1* (KD1 and KD2) and *Luciferase* (Luci). (c) qPCR data for *Dmrt1* and other pluripotency markers in transfected cells. Data represents the mean ± s.e.m. of two biological replicates (n = 2). (b) Phase images of transfected cells.

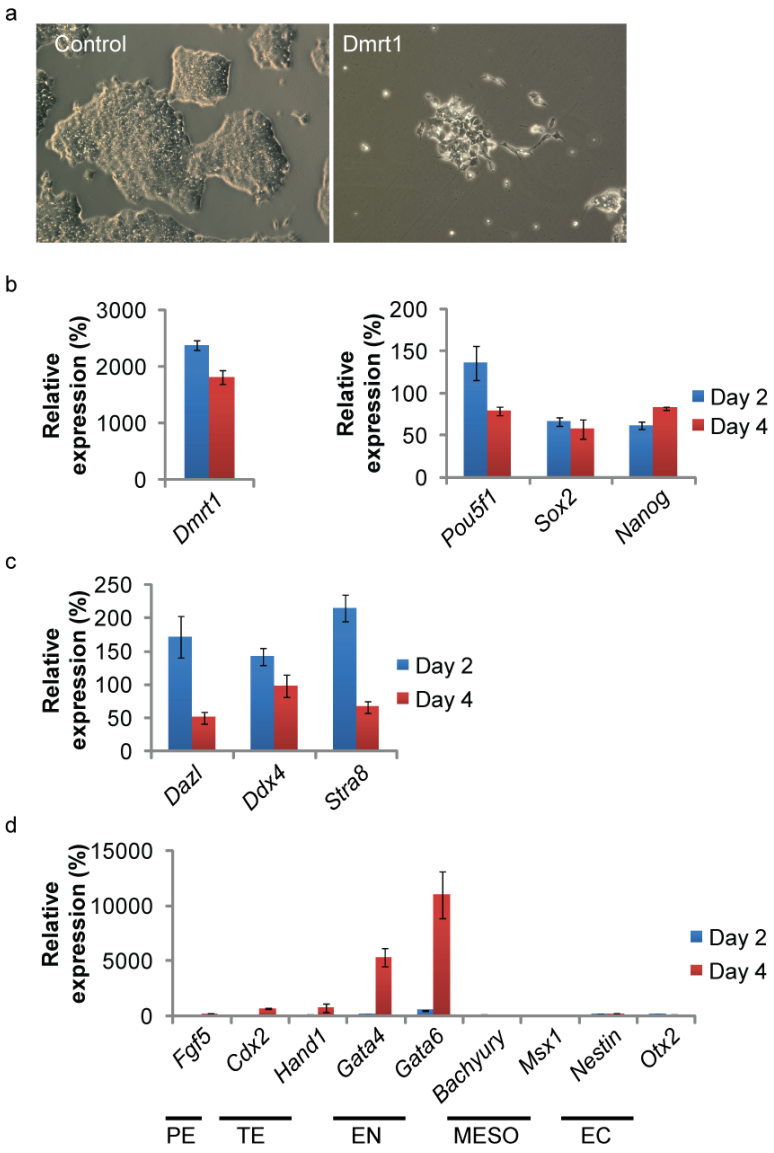

Figure 32: Overexpression of *Dmrt1* in E14 mESC caused differentiation. (a) Phase images of transfected cells after 3 days of selection. qPCR data for (b) *Dmrt1* and other pluripotency markers (c) germ cell markers and (d) other differentiation markers of primitive ectoderm (PE), trophectoderm (TE), endoderm (EN), mesoderm (MESO) and ectoderm (EC) in transfected cells on Day 2 and Day 4 of transfection. Data is normalized to empty vector control and represents the mean ± s.e.m. of three biological replicates (n = 3).

inability to sustain germ cell genes at the initial upregulated levels suggests that other critical germ cell regulators are required to induce and maintain a germ cell fate. It is also possible that the ectopic overexpression of *Dmrt1* to more than 15-fold endogenous level (Figure 32a) needs to be further fine-tuned to strike a balance between pluripotency gene repression and germ cell gene activation, in a manner more akin to *in vivo* germ cells at 13.5 d.p.c. Alternatively, a monolayer culture system might not be optimal for *in vitro* germ cell differentiation and the culture media probably requires optimization in order to support germ cell survival.

9.4.3. Nr5a2

Nr5a1 and Nr5a2, also known as steroidogenic factor 1 (Sf1) and liver receptor homolog 1 (Lrh1) respectively, are closely related members of the Nr5a or Ftz-F1 subfamily orphan nuclear receptors (Fayard et al., 2004). In the mouse embryo, Nr5a1 plays important functions in steroidogenic tissues such as the adrenals and gonads, which are both derived from the adrenogenital primordium (AGP) (Val et al., 2003). Within the gonad itself, while *Nr5a1* expression is not detectable in germ cells, its expression can be detected in the Sertoli and Leydig cells where it regulates steroidogenesis and hence sex differentiation (Hinshelwood et al., 2005). On the other hand, *Nr5a2* is expressed in endodermal-derived tissues such as the intestine, liver and pancreas and is involved in cholesterol metabolism and steroidogenesis (Fayard et al., 2004). In the embryonic gonad prior to E15.5, using *in situ* hybridization, *Nr5a2* was reported to be expressed in germ cells and pre-Sertoli cells within the testicular cords (Hinshelwood et al., 2005).

Based on our microarray data, *Nr5a2* is enriched in E11.5 and E13.5 germ cells as compared to non-germ cells of the gonad, but is downregulated by E15.5 (Figure 33a). qPCR validation results indicate that *Nr5a2* expression level in E13.5 male germ cells is 37-fold of the non-germ cell level (Figure 33b). Immunohistostaining further confirms that Nr5a2 is expressed in E13.5 male germ cells at the protein level (Figure 33c). In contrast, *Nr5a1* is expressed at higher levels in non-germ cells than E13.5 male germ cells (Figure 33b). Therefore, *Nr5a2* rather than *Nr5a1* is the family member that is more specifically expressed in germ cells.

Other than tissues of endodermal origin and the gonads, *Nr5a2* is also expressed in the ICM and epiblast in the early embryo (Gu et al., 2005). Interestingly, Nr5a2 is required to maintain Oct4 expression in the epiblast but not in the ICM, and was found to bind directly to the proximal enhancer and promoter of the *Pou5f1* gene (Gu et al., 2005). While *Nr5a2* knockout mESCs could be established and retain pluripotency as assayed by *in vitro* differentiation and *in vivo* teratoma formation (Gu et al., 2005; Wagner et al., 2010), Oct4 expression declined more rapidly in the knockout cells as compared to wildtype cells under differentiation-inducing conditions (Gu et al., 2005). Nr5a2 was also shown to enhance reprogramming efficiencies and is able to replace Oct4 as a reprogramming factor (Heng et al., 2010). More recently, Nr5a2 was identified as an essential factor for ground state pluripotency, in which mESCs can be maintained in the absence of LIF and BMP signals (Guo and Smith, 2010). By transfecting mESCs with shRNA constructs targeting *Nr5a2*, *Nr5a2* could be knockdown to less than 25% endogenous levels. After 3 days of selection, *Nr5a2* knockdown results in downregulation of *Nanog*, but

not *Pou5f1* and *Sox2* (Figure 34a), and the transfected mESCs exhibited an increased propensity to differentiate (Figure 34b).

Nr5a2-knockout results in embryonic lethality by 7.5 d.p.c. (Botrugno et al., 2004; Gu et al., 2005; Pare et al., 2004) so it is not possible to study its function in germ cells *in vivo*. To investigate if Nr5a2 has any potential functions in germ cells, I have undertaken an *in vitro* embryoid body-mediated method. The EOS reporter comprises three copies of the *Pou5f1* distal enhancer, and is an effective reporter used in iPSC isolation (Hotta et al., 2009) (Figure 35a). Undifferentiated mESCs show strong GFP expression, which is extinguished during differentiation (Figure 35b). Embryoid bodies (EBs) were formed and cultured in the absence of LIF over a period of 7 days, and samples were sorted every 2 days (Figure 36). Over the course of 7 days, there was a change in both GFP fluorescence as well as cell size, which is indicated by the forward scatter (FSC) parameter. Within a day of EB formation, there was a shift from cell population 1 to cell population 2 (Figure 37). Cell population 1 consists of mainly GFP-positive cells that are likely residual undifferentiated mESCs. In contrast to cell population 1, cell population 2 comprises a larger fraction of GFP-negative cells and a lower average FSC that implies a smaller cell size. Beginning from Day 3 onwards, a subset of cells within cell population 2 becomes visually distinguishable from the GFP-negative fraction (Figure 36). By Day 7, two subpopulations can be clearly seen from cell population 2. RNA was harvested from the GFP-positive sub-population of cell population 2 and analyzed for expression of pluripotency and germ cell markers. Expression of *Stella*, *Dazl*, *Tex19* and *Stra8* show an increasing trend up to Day 7 while the expression of other germ cell markers *Blimp1*, *Ddx4*, *Mov10l1* increased to a peak on Day 4 followed by a decline on Day 7 (Figure 38).

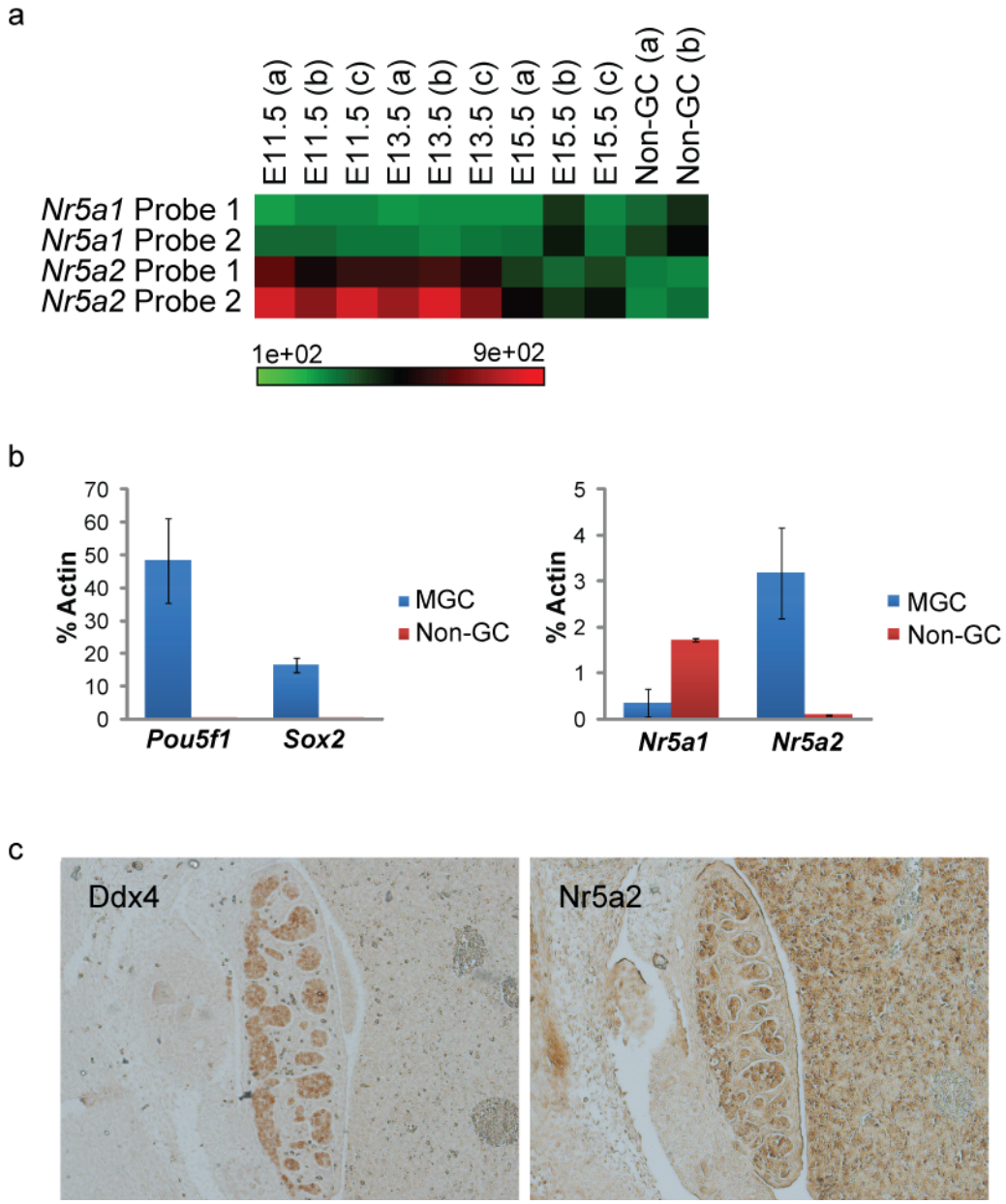

Figure 33: Nr5a2 is expressed in germ cells. (a) Microarray heatmaps for *Nr5a1* and *Nr5a2* in germ cells and non-germ cells. (b) qPCR results for pluripotency markers *Pou5f1* and *Sox2*, and *Nr5a1* and *Nr5a2* in E13.5 male germ cells (MGC) and non-germ cells (non-GC). Data represents the mean ± s.e.m. of two biological replicates (n = 2). (c) IHC staining of E13.5 male gonadal sections using anti-Ddx4 (post-migratory germ cell marker) and anti-Nr5a2 antibodies.

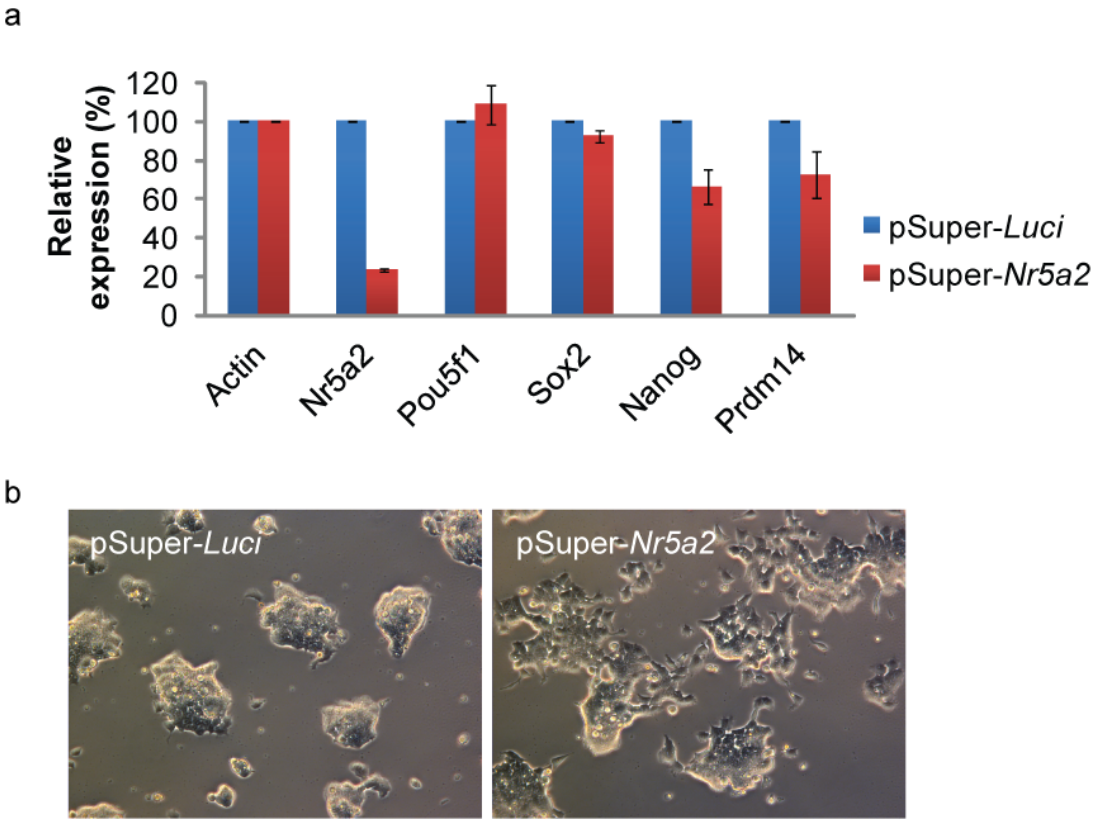

Figure 34: *Nr5a2* knockdown in E14 mESC. (a) qPCR data for *Nr5a2* and other pluripotency markers in transfected cells after 3 days of selection. Data represents the mean ± s.e.m. of three biological replicates (n = 3). (b) Phase images of transfected cells.

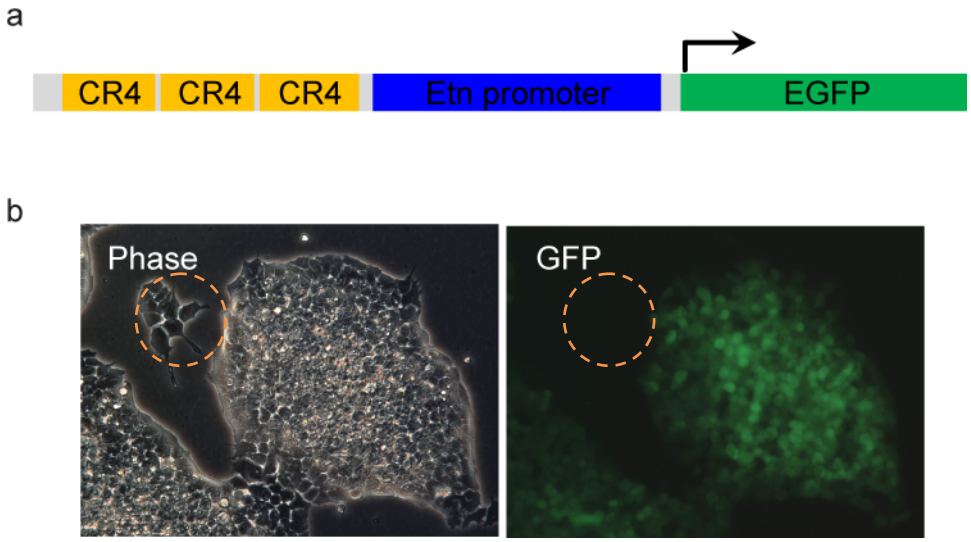

Figure 35: EOS-C(3+) reporter is specifically activated in undifferentiated mESC. (a) Schematic of EOS-C(3+) lentiviral vector containing a trimer of the *Pou5f1* distal enhancer (CR4). (b) EOS-C(3+) lentiviral-infected E14 mESCs. Dotted region shows differentiated cells that became GFP-negative.

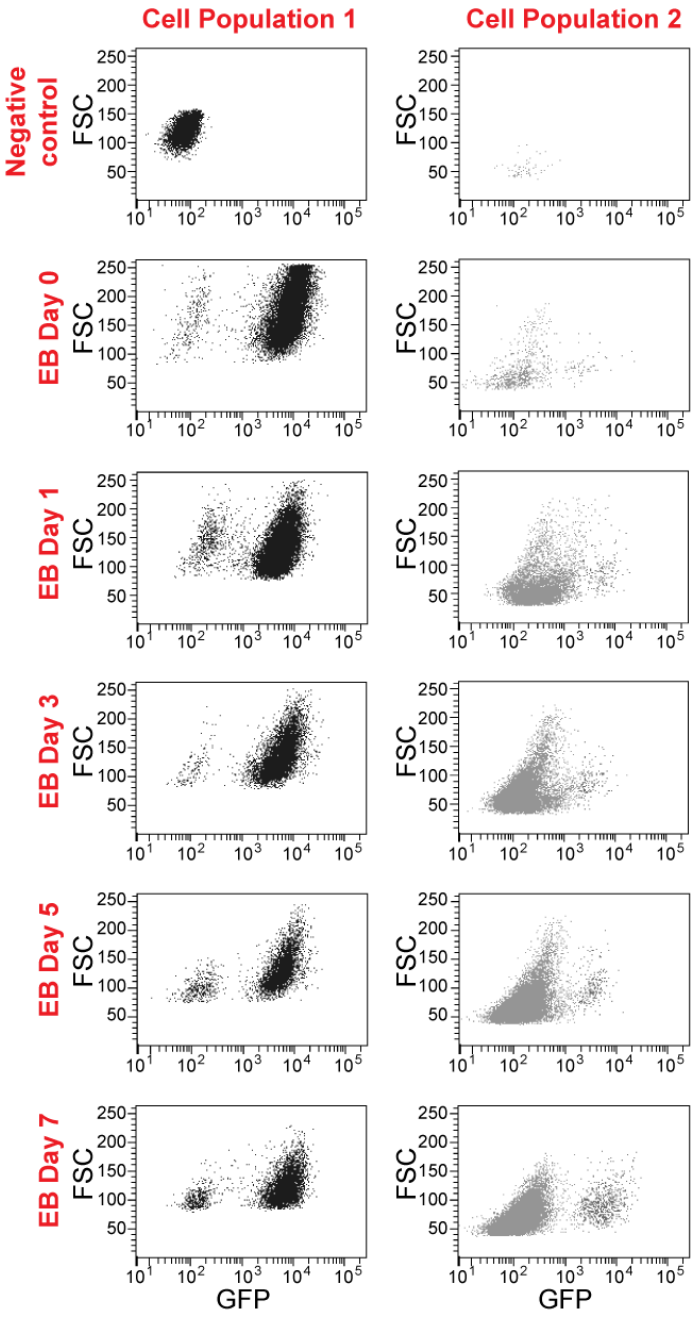

Figure 36: FACS profile of EOS-C(3+) embryoid bodies from Day 0 to Day 7.

FACS scatterplot of dissociated cells from EBs of different timepoints (Day 0 to Day 7). Y-axis is forward scatter (FSC) which is correlated to cell size and x-axis is GFP fluorescence intensity. Non-GFP-labeled mESC was used as negative control.

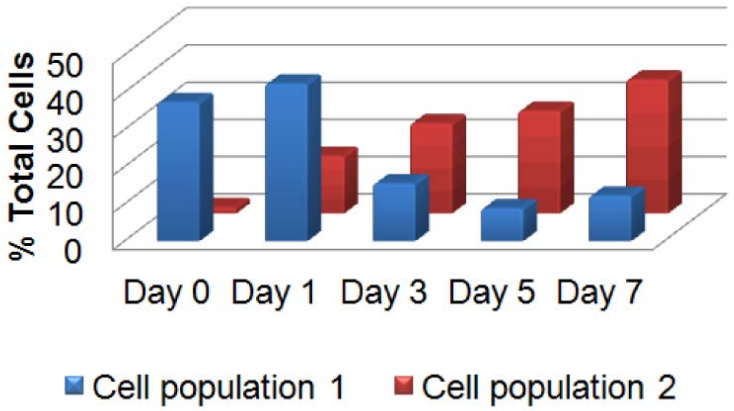

Figure 37: Cell population shift during EB differentiation.

The percentage of total cells that fall under cell population 1 (larger average FSC) declines while that of cell population 2 (smaller average FSC) increases from Day 0 to Day 7. This suggests a shift from cell population 1 to cell population 2, and a reduction of average cell size, as EBs are cultured over 7 days.

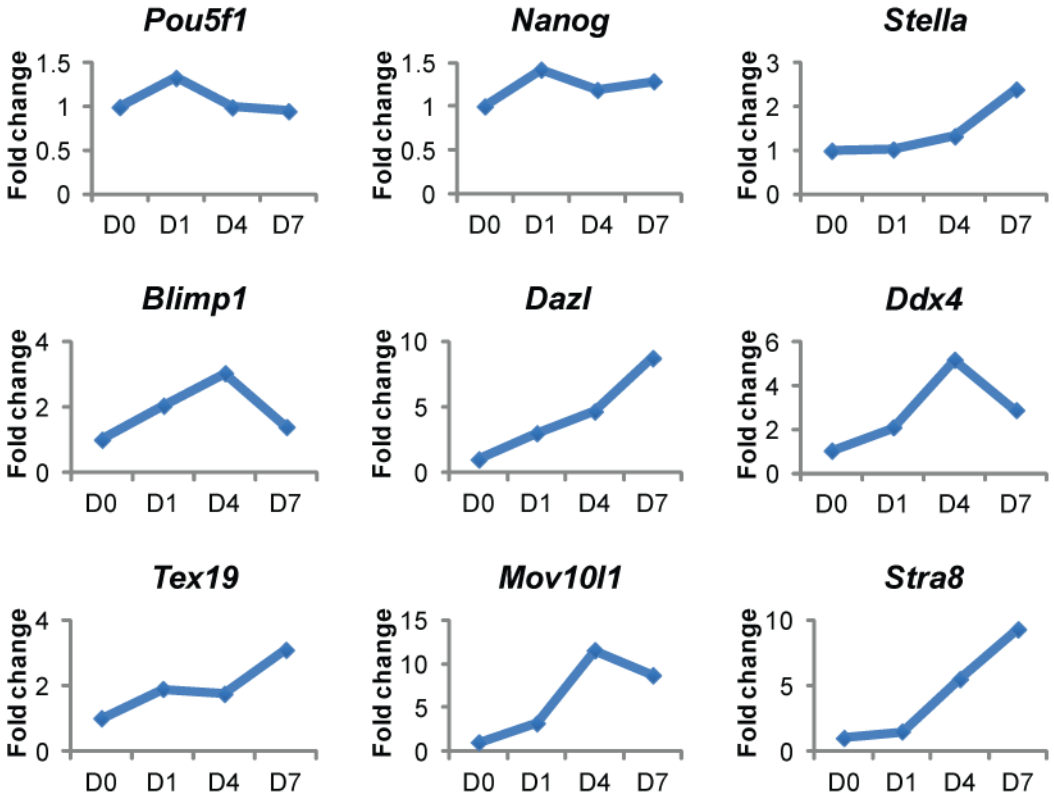

Figure 38: Enrichment of germ cell marker expression in EOS-C(3+) GFP-positive cells purified from EBs. Relative gene expression, as quantified by qPCR, is expressed as a fold change relative to respective gene expression level on Day 0 (D0).

Having established that the EB formation is a feasible method for *in vitro* germ cell differentiation, I wanted to test whether Nr5a2 affects germ cell differentiation *in vitro*. Mouse v6.4 ESCs were infected with EOS-C(3+) reporter lentiviruses along with either *Luciferase* or *Nr5a2* knockdown lentiviruses. Stable clonal cell lines were established after colony picking and several weeks of culture under selection pressure (Figure 39a). Concordant to published data (Gu et al., 2005; Guo and Smith, 2010), Nr5a2 is not essential for mESC maintenance. The stable *Nr5a2* knockdown cell line (EOS-Nr5a2) appeared morphologically similar to the control *Luci*ferase knockdown cell line (EOS-Luci), with dome-shaped colonies that are positive for the EOS-C(3+) EGFP reporter expression (Figure 39b). Both EOS-Luci and EOS-Nr5a2 knockdown cells were then subjected to EB formation for seven days. There was no visible difference in the EBs that were formed using the EOS-Luci and EOS-Nr5a2 cell lines, with uniformly bright GFP expression on Day 1 (Figure 40a) and heterogeneous GFP expression by Day 7 (Figure 40b). Furthermore, there was also no obvious distinction in the GFP intensity or pattern of the GFP-positive subset (from cell population 2) between EOS-Luci and EOS-Nr5a2 (Figure 40c).

Although the purified GFP-positive fraction is supposedly enriched for putative germ cells, as shown previously, it is difficult to exclude the possibility of the presence of undifferentiated ESCs which could mask the effect of *Nr5a2* knockdown. Therefore, the GFP-positive cells (from cell population 2) of both EOS-Luci EBs and EOS-Nr5a2 EBs were FACS-purified and cultured in EGC-conversion culture medium (EGM), which contains three essential growth factors namely, LIF, bFGF and SCF (Matsui et al., 1992; Resnick et al., 1992). In addition, retinoic acid (RA) was also included in the culture medium to differentiate any residual undifferentiated ESCs, as

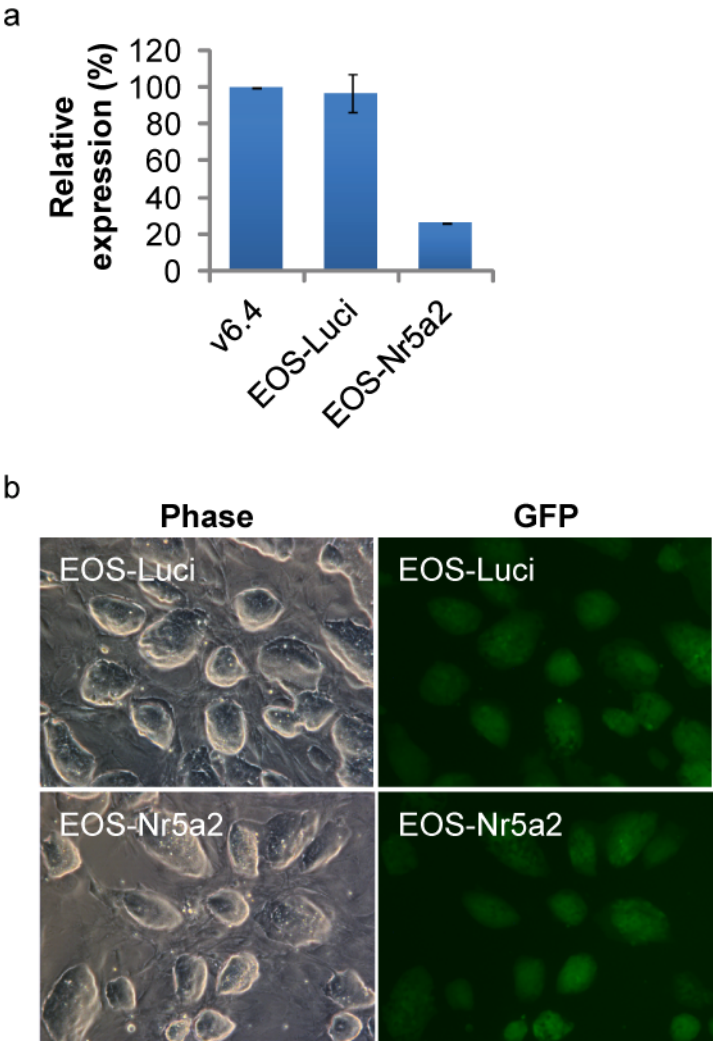

Figure 39: Establishment of stable EOS-Luci and EOS-Nr5a2 knockdown mESC lines. (a) qPCR data for *Nr5a2* levels in stable knockdown cell lines. Data represents the mean ± s.e.m. of two technical replicates (n = 2). (b) Phase and fluorescent images of EOS-Luci and EOS-Nr5a2 knockdown mESC lines.

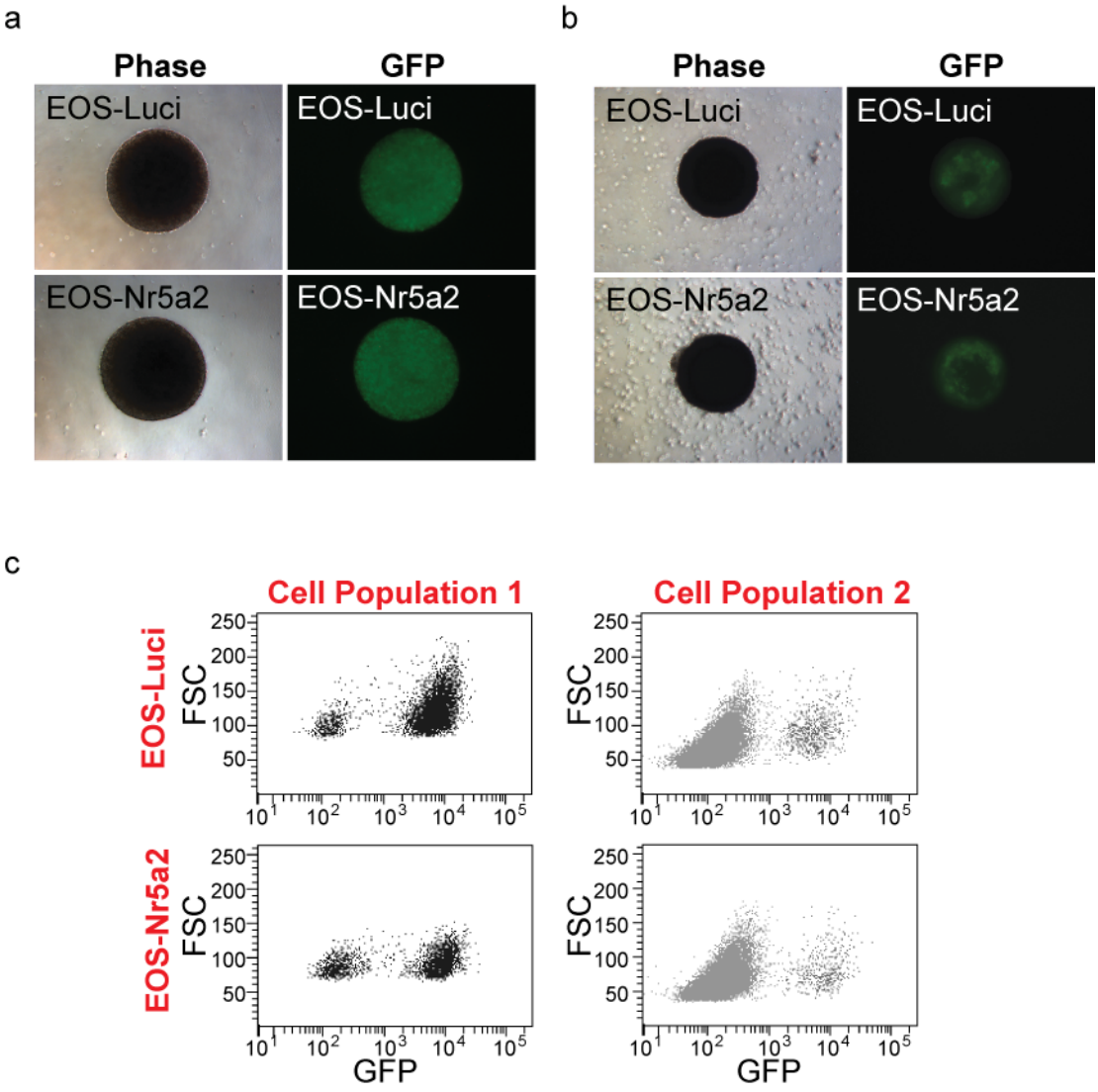

Figure 40: EB formation using stable EOS-Luci and EOS-Nr5a2 knockdown mESC lines. Phase and GFP images of embryoid bodies that were generated using EOS-Luci and EOS-Nr5a2 mESC on (a) day 1 and (b) day 7. (c) FACS profile of EOS-Luci and EOS-Nr5a2 embryoid bodies on Day 7.

well as to promote survival and proliferation of potential germ cells (West et al., 2006; West et al., 2009). EGC colonies were later detected by alkaline phosphatase (AP) staining after 5-7 days in culture. As compared to EOS-Luci, significantly fewer AP-positive colonies were observed for EOS-Nr5a2 (Figure 41c,d). As a control, GFP-positive cells were also purified from undifferentiated cell lines and grown in typical ESC medium (ESM) in the presence of LIF. Although EOS-Nr5a2 also had fewer AP-positive colonies than EOS-Luci in ESM conditions (Figure 41a-b), the reduction is much more evident for differentiated cells in EGM conditions (Figure 41c-d). Hence, the effect of *Nr5a2* knockdown on the number of AP-positive colonies from the germ-cell enriched sorted cells cannot be attributed to potential contamination by undifferentiated mESCs. This result then leads us to two possibilities: (1) Nr5a2 affects germ cell formation *in vitro* or (2) Nr5a2 affects the conversion efficiency of germ cells to EGC. To investigate further, RNA was extracted from the "germ cell-enriched" GFP-positive cells of Day 7 EBs and, pluripotency and germ cell marker expression was examined using realtime qPCR. While pluripotency factors such as *Pou5f1*, *Sox2* and *Nanog* were expressed at slightly lower levels in the *Nr5a2* knockdown GFP-positive cells as compared to *Luciferase* knockdown control, there was a greater reduction in the expression of germ cell markers, especially for later germ cell markers such as *Dazl*, *Ddx4*, *Mov10l1*, *Stra8* and *Sycp3* (Figure 41e). Altogether, our data implies that Nr5a2 may be important for germ cell differentiation.

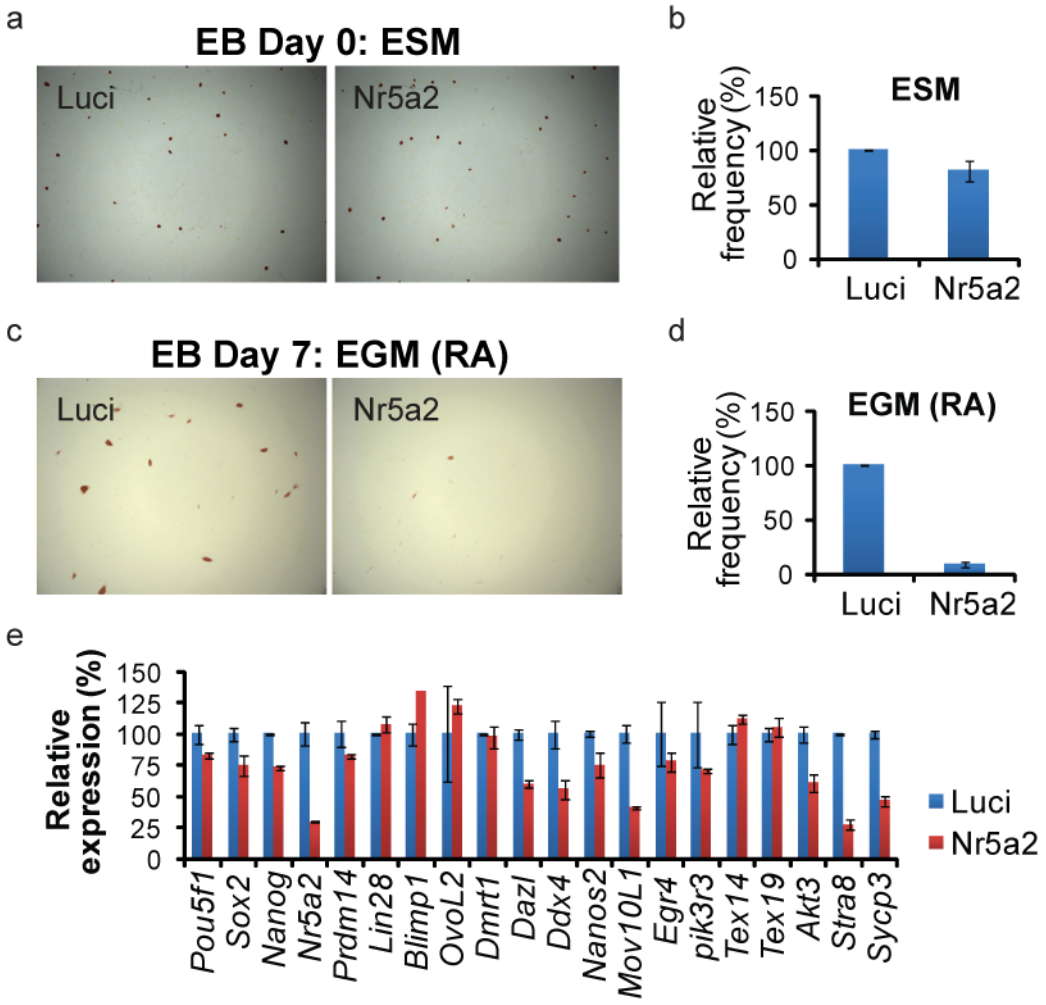

Figure 41: Nr5a2 plays a role in germ cell formation *in vitro*. Images and quantification of AP-positive colonies that form when GFP-positive EOS-Luci and EOS-Nr5a2 cells purified from (a,b) Day 0 and (c,d) Day 7 EBs were cultured in ESM and EGM (RA) respectively. Quantitation of AP-positive colonies is expressed as a ratio of the respective *Luciferase* knockdown control frequencies. Data represents the mean ± s.e.m. of two biological replicates (n = 2) for ESM and three biological replicates (n = 3) for EGM (RA). (e) Realtime qPCR data for pluripotency and germ cell markers in GFP-positive cells that were purified from Day 7 EBs. Expression changes were normalized to *Luciferase* knockdown control.

10. **Conclusion**

Using both ESCs and FACS-purified germ cells, I have demonstrated that ChIP-Seq could be achieved from minute starting amounts of DNA or small cell populations. The additional WGA amplification step and subsequent procedures to excise PCR adaptors do not adversely affect ChIP-Seq results. Genome-wide data of ChIP duplicates also suggests reproducibility of the small-scale ChIP procedures. However, good quality ChIP-Seq results still depend on purification of sufficient cell numbers (at least 50,000 cells) and the availability of good ChIP-grade antibodies. While I have generated histone ChIP-Seq libraries, it remains to be tested whether the same procedure is adequate for transcription factor ChIP-Seq. Given that DNA yield and ChIP enrichment tend to be less for transcription factor ChIP, further optimization might be required.

In this project, I have optimized and performed genome-wide techniques to study *in vivo* germ cells. More specifically, the combination of microarray and ChIP-Seq provides a powerful method for mining useful information on transcriptional and epigenetic regulation. Microarray analyses suggest that E13.5 germ cells are undergoing a transitional period whereby germ cells are driven to differentiate and "pluipotential" is lost. Early-changing genes may be important in mediating this switch at E13.5. For example, the upregulation of $p16^{INK4A}$, which was previously reported as a barrier to iPSC reprogramming (Li et al., 2009), may explain why germ cells cannot revert back to pluripotent EGCs after E12.5 (Labosky et al., 1994), even though many pluripotency-associated genes are transcriptionally downregulated only after E13.5. More experiments would be needed to further test this tempting but speculative conjecture.

Another area of interest is to investigate the similarities and differences between germ cells, ESCs and other differentiated cell types, and identify transcription factors that contribute to the germ cell identity. The genome-wide maps of H3K4me3, H3K27me3, H3K27ac and H2BK20ac in E13.5 male germ cells provide a means to determine the activity status of regulatory elements such as promoters and enhancers. Taking advantage of publicly available histone modification data for other cell types, it is possible to identify promoters and enhancers that are specifically activated in germ cells. The enrichment of germ cell related processes and mouse phenotypes also render some confidence to biological relevance of these identified regulatory elements.

Going a step further, potential germ cell regulators were identified by undertaking several strategies that involve motif analyses and making use of available transcription factor ChIP-Seq data. Two such factors, Dmrt1 and Nr5a2, have been selected for further in-depth studies. Expression has been validated on the transcript and protein level, and both factors are shown to be enriched in germ cells. Although Dmrt1 has already been implicated in germ cell development, this study reveals that the Dmrt motif is enriched at not only promoters but also active enhancers in E13.5 male germ cells. Therefore, future ChIP-Seq data of Dmrt1 could yield more insight into its downstream targets and enhance our understanding of its regulatory functions in germ cells. On the other hand, the discovery of a potential role of Nr5a2 in germ cells is novel and not previously reported. Given that *Nr5a2* knockout causes embryonic lethality (Botrugno et al., 2004; Gu et al., 2005; Pare et al., 2004) prior to the developmental time frame of this study, I have taken an alternative *in vitro* approach to show that Nr5a2 plays a role in germ cell differentiation. However, to be

absolutely certain about this conclusion, it is essential to demonstrate the role of Nr5a2 *in vivo* via the generation of a germ cell-specific conditional knockout.

Lastly, this study demonstrates the feasibility of generating ChIP-Seq data from small cell populations and opens the possibility to ChIP other types of rare cells.

11. **Bibliography**

Aasen, T., Raya, A., Barrero, M.J., Garreta, E., Consiglio, A., Gonzalez, F., Vassena, R., Bilic, J., Pekarik, V., Tiscornia, G., *et al.* (2008). Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. Nat Biotechnol *26*, 1276-1284.

Aoi, T., Yae, K., Nakagawa, M., Ichisaka, T., Okita, K., Takahashi, K., Chiba, T., and Yamanaka, S. (2008). Generation of pluripotent stem cells from adult mouse liver and stomach cells. Science *321*, 699-702.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev *17*, 126-140.

Ayad, N.G., Rankin, S., Murakami, M., Jebanathirajah, J., Gygi, S., and Kirschner, M.W. (2003). Tome-1, a Trigger of Mitotic Entry, Is Degraded during G1 via the APC. Cell *113*, 101-113.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., *et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315-326.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

Botrugno, O.A., Fayard, E., Annicotte, J.-S.b., Haby, C.l., Brennan, T., Wendling, O., Tanaka, T., Kodama, T., Thomas, W., Auwerx, J., *et al.* (2004). Synergy between LRH-1 and B-Catenin Induces G1 Cyclin-Mediated Cell Proliferation. Molecular Cell *15*, 499-509.

Bourc'his, D.b., Xu, G.-L., Lin, C.-S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the Establishment of Maternal Genomic Imprints. Science *294*, 2536-2539.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255-256.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., *et al.* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.
Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development *132*, 885-896.
Cereghini, S. (1996). Liver-enriched transcription factors and hepatocyte differentiation. The FASEB Journal *10*, 267-282.
Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230-1234.
Chen, X., Fang, F., Liou, Y.-C., and Ng, H.-H. (2008a). Zfp143 Regulates Nanog Through Modulation of Oct4 Binding. STEM CELLS *26*, 2759-2767.
Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., *et al.* (2008b). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. Cell *133*, 1106-1117.
Chinwalla, A.T. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.
Chong, J.-L., Wenzel, P.L., Saenz-Robles, M.T., Nair, V., Ferrey, A., Hagan, J.P., Gomez, Y.M., Sharma, N., Chen, H.-Z., Ouseph, M., *et al.* (2009). E2f1-3 switch from activators in progenitor cells to repressors in differentiating cells. Nature *462*, 930-934.
Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev *22*, 746-755.
Cowan, C.A., Atienza, J., Melton, D.A., and Eggan, K. (2005). Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science *309*, 1369-1373.
Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences *107*, 21931-21936.
da Silva, S.M., Hacker, A., Harley, V., Goodfellow, P., Swain, A., and Lovell-Badge, R. (1996). Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. Nat Genet *14*, 62-68.
De La Fuente, R., Baumann, C., Fan, T., Schmidtmann, A., Dobrinski, I., and Muegge, K. (2006). Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. Nat Cell Biol *8*, 1448-1454.
Dejosez, M., Krumenacker, J.S., Zitur, L.J., Passeri, M., Chu, L.-F., Songyang, Z., Thomson, J.A., and Zwaka, T.P. (2008). Ronin Is Essential for Embryogenesis and the Pluripotency of Mouse Embryonic Stem Cells. Cell *133*, 1162-1174.
Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M.M., Heninger, A.-K., Vries, I.d., Kittler, R., Junqueira, M., Shevchenko, A., Schulz, H., *et al.* (2009). A Genome-Scale RNAi Screen for Oct4 Modulators Defines a Role of the Paf1 Complex for Embryonic Stem Cell Identity. Cell Stem Cell *4*, 403-415.

Dolci, S., Williams, D.E., Ernst, M.K., Resnick, J.L., Brannan, C.I., Lock, L.F., Lyman, S.D., Boswell, H.S., and Donovan, P.J. (1991). Requirement for mast cell growth factor for primordial germ cell survival in culture. Nature *352*, 809-811.
Donovan, P.J., Stott, D., Cairns, L.A., Heasman, J., and Wylie, C.C. (1986). Migratory and postmigratory mouse primordial germ cells behave differently in culture. Cell *44*, 831-838.
Durcova-Hills, G., Tang, F., Doody, G., Tooze, R., and Surani, M.A. (2008). Reprogramming Primordial Germ Cells into Pluripotent Stem Cells. PLoS ONE *3*, e3531.
Endoh, M., Endo, T.A., Endoh, T., Fujimura, Y.-i., Ohara, O., Toyoda, T., Otte, A.P., Okano, M., Brockdorff, N., Vidal, M., *et al.* (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development *135*, 1513-1524.
Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.
Farthing, C.R., Ficz, G., Ng, R.K., Chan, C.F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet *4*, e1000116.
Fayard, E., Auwerx, J., and Schoonjans, K. (2004). LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis. Trends in Cell Biology *14*, 250-260.
Feng, B., Jiang, J., Kraus, P., Ng, J.-H., Heng, J.-C.D., Chan, Y.-S., Yaw, L.-P., Zhang, W., Loh, Y.-H., Han, J., *et al.* (2009a). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat Cell Biol *11*, 197-203.
Feng, B., Ng, J.-H., Heng, J.-C.D., and Ng, H.-H. (2009b). Molecules that Promote or Enhance Reprogramming of Somatic Cells to Induced Pluripotent Stem Cells. Cell Stem Cell *4*, 301-312.
Ferri, A.L.M., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., Ottolenghi, S., Pandolfi, P.P., Sala, M., DeBiasi, S., *et al.* (2004). Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. Development *131*, 3805-3819.
Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.-K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, Esteban O., *et al.* (2011). An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming. Cell *147*, 132-146.
Godin, I., Deed, R., Cooke, J., Zsebo, K., Dexter, M., and Wylie, C.C. (1991). Effects of the steel gene product on mouse primordial germ cells in culture. Nature *352*, 807-809.
Gu, P., Goodwin, B., Chung, A.C.K., Xu, X., Wheeler, D.A., Price, R.R., Galardi, C., Peng, L., Latour, A.M., Koller, B.H., *et al.* (2005). Orphan Nuclear Receptor LRH-1 Is Required To Maintain Oct4 Expression at the Epiblast Stage of Embryonic Development. Molecular and Cellular Biology *25*, 3492-3505.
Guo, G., and Smith, A. (2010). A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. Development *137*, 3185-3192.

Hanna, J., Markoulaki, S., Schorderet, P., Carey, B.W., Beard, C., Wernig, M., Creyghton, Menno P., Steine, E.J., Cassady, J.P., Foreman, R., *et al.* (2008). Direct Reprogramming of Terminally Differentiated Mature B Lymphocytes to Pluripotency. Cell *133*, 250-264.
Hannenhalli, S., and Kaestner, K.H. (2009). The evolution of Fox genes and their role in development and disease. Nat Rev Genet *10*, 233-240.
Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development *129*, 1983-1993.
Hayashi, K., Yoshida, K., and Matsui, Y. (2005). A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. Nature *438*, 374-378.
Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., *et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.
Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.
Heng, J.-C.D., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J.-H., Orlov, Y.L., Huss, M., Yang, L., Lufkin, T., *et al.* (2010). The Nuclear Receptor Nr5a2 Can Replace Oct4 in the Reprogramming of Murine Somatic Cells to Pluripotent Cells. Cell Stem Cell *6*, 167-174.
Hinshelwood, M.M., Shelton, J.M., Richardson, J.A., and Mendelson, C.R. (2005). Temporal and spatial expression of liver receptor homologue-1 (LRH-1) during embryogenesis suggests a potential role in gonadal development. Dev Dyn *234*, 159-168.
Hogarth, C.A., Mitchell, D., Small, C., and Griswold, M. (2010). EGR4 displays both a cell- and intracellular-specific localization pattern in the developing murine testis. Developmental Dynamics *239*, 3106-3114.
Hong, C.-S., Park, B.-Y., and Saint-Jeannet, J.-P. (2007). The function of Dmrt genes in vertebrate development: It is not just about sex. Developmental Biology *310*, 1-9.
Hotta, A., Cheung, A.Y.L., Farra, N., Vijayaragavan, K., Seguin, C.A., Draper, J.S., Pasceri, P., Maksakova, I.A., Mager, D.L., Rossant, J., *et al.* (2009). Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. Nat Meth *6*, 370-376.
Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes & Development *23*, 837-848.
Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protocols *4*, 44-57.
Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research *37*, 1-13.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature *442*, 533-538.
Jaenisch, R., and Young, R. (2008). Stem Cells, the Molecular Circuitry of Pluripotency and Nuclear Reprogramming. Cell *132*, 567-582.
Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol *10*, 353-360.
Jopling, C., Boue, S., and Belmonte, J.C.I. (2011). Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. Nat Rev Mol Cell Biol *12*, 79-89.
Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., *et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.
Kanetsky, P.A., Mitra, N., Vardhanabhuti, S., Vaughn, D.J., Li, M., Ciosek, S.L., Letrero, R., D'Andrea, K., Vaddi, M., Doody, D.R., *et al.* (2011). A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility. Human Molecular Genetics *20*, 3109-3117.
Kehler, J., Tolkunova, E., Koschorz, B., Pesce, M., Gentile, L., Boiani, M., Lomeli, H., Nagy, A., McLaughlin, K.J., Scholer, H.R., *et al.* (2004). Oct4 is required for primordial germ cell survival. EMBO Rep *5*, 1078-1083.
Kim, D., Kim, C.-H., Moon, J.-I., Chung, Y.-G., Chang, M.-Y., Han, B.-S., Ko, S., Yang, E., Cha, K.Y., Lanza, R., *et al.* (2009a). Generation of Human Induced Pluripotent Stem Cells by Direct Delivery of Reprogramming Proteins. Cell Stem Cell *4*, 472-476.
Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049-1061.
Kim, J.B., Sebastiano, V., Wu, G., Araúzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D., *et al.* (2009b). Oct4-Induced Pluripotency in Adult Neural Stem Cells. Cell *136*, 411-419.
Kim, S., Bardwell, V.J., and Zarkower, D. (2007a). Cell type-autonomous and non-autonomous requirements for Dmrt1 in postnatal testis differentiation. Developmental Biology *307*, 314-327.
Kim, S., Kettlewell, J.R., Anderson, R.C., Bardwell, V.J., and Zarkower, D. (2003). Sexually dimorphic expression of multiple doublesex-related genes in the embryonic mouse gonad. Gene Expression Patterns *3*, 77-82.
Kim, S., Namekawa, S.H., Niswander, L.M., Ward, J.O., Lee, J.T., Bardwell, V.J., and Zarkower, D. (2007b). A Mammal-Specific Doublesex Homolog Associates with Male Sex Chromatin and Is Required for Male Meiosis. PLoS Genet *3*, e62.
Kleinsmith, L.J., and Pierce, G.B. (1964). Multipotentiality of Single Embryonal Carcinoma Cells. Cancer Research *24*, 1544-1551.
Kocer, A., Reichmann, J., Best, D., and Adams, I.R. (2009). Germ cell sex determination in mammals. Molecular Human Reproduction *15*, 205-213.
Kooistra, S.M., van den Boom, V., Thummer, R.P., Johannes, F., Wardenaar, R., Tesson, B.M., Veenhoff, L.M., Fusetti, F., O'Neill, L.P., Turner, B.M., *et al.* (2010).

Undifferentiated Embryonic Cell Transcription Factor 1 Regulates ESC Chromatin Organization and Gene Expression. STEM CELLS *28*, 1703-1714.
Krentz, A.D., Murphy, M.W., Kim, S., Cook, M.S., Capel, B., Zhu, R., Matin, A., Sarver, A.L., Parker, K.L., Griswold, M.D., *et al.* (2009). The DM domain protein DMRT1 is a dose-sensitive regulator of fetal germ cell proliferation and pluripotency. Proc Natl Acad Sci U S A *106*, 22323-22328.
Krentz, A.D., Murphy, M.W., Sarver, A.L., Griswold, M.D., Bardwell, V.J., and Zarkower, D. (2011). DMRT1 promotes oogenesis by transcriptional activation of Stra8 in the mammalian fetal ovary. Dev Biol *356*, 63-70.
Labosky, P.A., Barlow, D.P., and Hogan, B.L. (1994). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. Development *120*, 3197-3204.
Lai, A., Kennedy, B.K., Barbie, D.A., Bertos, N.R., Yang, X.J., Theberge, M.-C., Tsai, S.-C., Seto, E., Zhang, Y., Kuzmichev, A., *et al.* (2001). RBP1 Recruits the mSIN3-Histone Deacetylase Complex to the Pocket of Retinoblastoma Tumor Suppressor Family Proteins Found in Limited Discrete Regions of the Nucleus at Growth Arrest. Mol Cell Biol *21*, 2918-2932.
Lei, N., Hornbaker, K.I., Rice, D.A., Karpova, T., Agbor, V.A., and Heckert, L.L. (2007). Sex-Specific Differences in Mouse DMRT1 Expression Are Both Cell Type- and Stage-Dependent During Gonad Development. Biology of Reproduction *77*, 466-475.
Li, H., Collado, M., Villasante, A., Strati, K., Ortega, S., Canamero, M., Blasco, M.A., and Serrano, M. (2009). The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature *460*, 1136-1139.
Lim, L.S., Loh, Y.-H., Zhang, W., Li, Y., Chen, X., Wang, Y., Bakre, M., Ng, H.-H., and Stanton, L.W. (2007). Zic3 Is Required for Maintenance of Pluripotency in Embryonic Stem Cells. Molecular Biology of the Cell *18*, 1348-1358.
Lin, C.-H., Jackson, A.L., Guo, J., Linsley, P.S., and Eisenman, R.N. (2009). Myc-regulated microRNAs attenuate embryonic stem cell differentiation. EMBO J *28*, 3157-3170.
Lin, Y., and Page, D.C. (2005). Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice. Developmental Biology *288*, 309-316.
Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440.
Loh, Y.-H., Zhang, W., Chen, X., George, J., and Ng, H.-H. (2007). Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. Genes & Development *21*, 2545-2557.
Maroulakou, I.G., and Bowe, D.B. (2000). Expression and function of Ets transcription factors in mammalian development: a regulatory network. Oncogene *19*, 6432-6442.
Martin, G.R. (1981). Isolation of a Pluripotent Cell Line from Early Mouse Embryos Cultured in Medium Conditioned by Teratocarcinoma Stem Cells. Proceedings of the National Academy of Sciences of the United States of America *78*, 7634-7638.

Matson, C.K., Murphy, M.W., Griswold, M.D., Yoshida, S., Bardwell, V.J., and Zarkower, D. (2010). The mammalian doublesex homolog DMRT1 is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. Dev Cell *19*, 612-624.
Matsui, Y., Toksoz, D., Nishikawa, S., Nishikawa, S.-I., Williams, D., Zsebo, K., and Hogan, B.L.M. (1991). Effect of Steel factor and leukaemia inhibitory factor on murine primordial germ cells in culture. Nature *353*, 750-752.
Matsui, Y., Zsebo, K., and Hogan, B.L.M. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell *70*, 841-847.
Mattout, A., and Meshorer, E. (2010). Chromatin plasticity and genome organization in pluripotent embryonic stem cells. Curr Opin Cell Biol *22*, 334-341.
Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.
Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol *7*, 540-546.
Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49-55.
Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.
Moreno, S.G., Attali, M., Allemand, I., Messiaen, S., Fouchet, P., Coffigny, H., Romeo, P.H., and Habert, R. (2010). TGFbeta signaling in male germ cells regulates gonocyte quiescence and fertility in mice. Dev Biol *342*, 74-84.
Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotech *26*, 101-106.
Nakata, Y., Shetzline, S., Sakashita, C., Kalota, A., Rallapalli, R., Rudnick, S.I., Zhang, Y., Emerson, S.G., and Gewirtz, A.M. (2007). c-Myb Contributes to G2/M Cell Cycle Transition in Human Hematopoietic Cells by Direct Regulation of Cyclin B1 Expression. Mol Cell Biol *27*, 2048-2058.
Ng, H.-H., and Surani, M.A. (2011). The transcriptional and signalling networks of pluripotency. Nat Cell Biol *13*, 490-496.
Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.
Niwa, H., Miyazaki, J.-i., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372-376.
Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A.*, et al.* (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature *436*, 207-213.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease. Cell Stem Cell *6*, 382-395.
Pare, J.-F.o., Malenfant, D., Courtemanche, C., Jacob-Wagner, M.v., Roy, S., Allard, D., and BÃ©langer, L. (2004). The Fetoprotein Transcription Factor (FTF) Gene Is Essential to Embryogenesis and Cholesterol Homeostasis and Is Regulated by a DR4 Element. Journal of Biological Chemistry *279*, 21206-21216.
Park, I.H., Lerou, P.H., Zhao, R., Huo, H., and Daley, G.Q. (2008). Generation of human-induced pluripotent stem cells. Nat Protoc *3*, 1180-1186.
Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet *10*, 669-680.
Pesce, M., Farrace, M.G., Piacentini, M., Dolci, S., and De Felici, M. (1993). Stem cell factor and leukemia inhibitory factor promote primordial germ cell survival by suppressing programmed cell death (apoptosis). Development *118*, 1089-1094.
Pesce, M., Wang, X., Wolgemuth, D.J., and SchÃ¶ler, H.R. (1998). Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. Mechanisms of Development *71*, 89-98.
Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.
Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.
Raymond, C.S., Kettlewell, J.R., Hirsch, B., Bardwell, V.J., and Zarkower, D. (1999). Expression of Dmrt1 in the Genital Ridge of Mouse and Chicken Embryos Suggests a Role in Vertebrate Sexual Development. Developmental Biology *215*, 208-220.
Raymond, C.S., Murphy, M.W., O'Sullivan, M.G., Bardwell, V.J., and Zarkower, D. (2000). Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. Genes & Development *14*, 2587-2595.
Redmer, T., Diecke, S., Grigoryan, T., Quiroga-Negreira, A., Birchmeier, W., and Besser, D. (2011). E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. EMBO Rep *12*, 720-726.
Reik, W., Dean, W., and Walter, J.r. (2001). Epigenetic Reprogramming in Mammalian Development. Science *293*, 1089-1093.
Resnick, J.L., Bixler, L.S., Cheng, L., and Donovan, P.J. (1992). Long-term proliferation of mouse primordial germ cells in culture. Nature *359*, 550-551.
Richardson, B.E., and Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: strategies from different organisms. Nat Rev Mol Cell Biol *11*, 37-49.
Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet *9*, 129-140.
Schmidt, M., Huber, L., Majdazari, A., Schutz, G., Williams, T., and Rohrer, H. (2011). The transcription factors AP-2Beta and AP-2alpha are required for survival of sympathetic progenitors and differentiated sympathetic neurons. Developmental Biology *355*, 89-100.
Scholer, H.R., Dressler, G.R., Balling, R., Rohdewohld, H., and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. EMBO J *9*, 2185-2195.

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Developmental Biology *278*, 440-458.
Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y., and Saitou, M. (2007). Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development *134*, 2627-2638.
Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., *et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Research *15*, 1034-1050.
Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722-737.
Singh, A.M., and Dalton, S. (2009). The Cell Cycle and Myc Intersect with Mechanisms that Regulate Pluripotency and Reprogramming. Cell Stem Cell *5*, 141-149.
Smith, K.N., Singh, A.M., and Dalton, S. (2010). Myc Represses Primitive Endoderm Differentiation in Pluripotent Stem Cells. Cell Stem Cell *7*, 343-354.
Solter, D., Skreb, N., and Damjanov, I. (1970). Extrauterine growth of mouse egg-cylinders results in malignant teratoma. Nature *227*, 503-504.
Stadtfeld, M., Brennand, K., and Hochedlinger, K. (2008). Reprogramming of pancreatic beta cells into induced pluripotent stem cells. Curr Biol *18*, 890-894.
Stevens, L.C. (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. Developmental Biology *21*, 364-382.
Stevens, L.C., and Little, C.C. (1954). Spontaneous Testicular Teratomas in an Inbred Strain of Mice. Proceedings of the National Academy of Sciences of the United States of America *40*, 1080-1087.
Stewart, C.L., Gadi, I., and Bhatt, H. (1994). Stem Cells from Primordial Germ Cells Can Reenter the Germ Line. Developmental Biology *161*, 626-628.
Suzuki, A., and Saga, Y. (2008). Nanos2 suppresses meiosis and promotes male germ cell differentiation. Genes & Development *22*, 430-435.
Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr Biol *11*, 1553-1558.
Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.
Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663-676.
Takigawa, Y., Hata, K., Muramatsu, S., Amano, K., Ono, K., Wakabayashi, M., Matsuda, A., Takada, K., Nishimura, R., and Yoneda, T. (2010). The transcription factor Znf219 regulates chondrocyte differentiation by assembling a transcription factory with Sox9. Journal of Cell Science *123*, 3780-3788.

Tam, P.P.L., and Snow, M.H.L. (1981). Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos. Journal of Embryology and Experimental Morphology *64*, 133-147.
Tam, W.-L., Lim, C.Y., Han, J., Zhang, J., Ang, Y.-S., Ng, H.-H., Yang, H., and Lim, B. (2008). T-Cell Factor 3 Regulates Embryonic Stem Cell Pluripotency and Self-Renewal by the Transcriptional Control of Multiple Lineage Pathways. STEM CELLS *26*, 2019-2031.
Tamashiro, K.L., Wakayama, T., Akutsu, H., Yamazaki, Y., Lachey, J.L., Wortman, M.D., Seeley, R.J., D'Alessio, D.A., Woods, S.C., Yanagimachi, R.*, et al.* (2002). Cloned mice have an obese phenotype not transmitted to their offspring. Nat Med *8*, 262-267.
Tanaka, S.S., Toyooka, Y., Akasu, R., Katoh-Fukui, Y., Nakahara, Y., Suzuki, R., Yokoyama, M., and Noce, T. (2000). The mouse homolog of Drosophila Vasa is required for the development of male germ cells. Genes & Development *14*, 841-853.
Tao, Y., Xi, S., Shan, J., Maunakea, A., Che, A., Briones, V., Lee, E.Y., Geiman, T., Huang, J., Stephens, R.*, et al.* (2011). Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. Proceedings of the National Academy of Sciences *108*, 5626-5631.
Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.
Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic Stem Cell Lines Derived from Human Blastocysts. Science *282*, 1145-1147.
Tourtellotte, W.G., Nagarajan, R., Auyeung, A., Mueller, C., and Milbrandt, J. (1999). Infertility associated with incomplete spermatogenic arrest and oligozoospermia in Egr4-deficient mice. Development *126*, 5061-5071.
Tourtellotte, W.G., Nagarajan, R., Bartke, A., and Milbrandt, J. (2000). Functional Compensation by Egr4 inEgr1-Dependent Luteinizing Hormone Regulation and Leydig Cell Steroidogenesis. Molecular and Cellular Biology *20*, 5261-5268.
Turnbull, C., Rapley, E.A., Seal, S., Pernet, D., Renwick, A., Hughes, D., Ricketts, M., Linger, R., Nsengimana, J., Deloukas, P.*, et al.* (2010). Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. Nat Genet *42*, 604-607.
Val, P., Lefrancois-Martinez, A.-M., Veyssiere, G., and Martinez, A. (2003). SF-1 a key player in the development and differentiation of steroidogenic tissues. Nuclear Receptor *1*, 8.
van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. Cell Stem Cell *6*, 369-381.
Veith, A.-M., Klattig, J.r., Dettai, A., Schmidt, C., Englert, C., and Volff, J.-N. (2006). Male-biased expression of X-chromosomal DM domain-less Dmrt8 genes in the mouse. Genomics *88*, 185-195.
Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F.*, et al.* (2009a). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009b). Genomic views of distant-acting enhancers. Nature *461*, 199-205.
Wagner, R.T., Xu, X., Yi, F., Merrill, B.J., and Cooney, A.J. (2010). Canonical Wnt/beta-catenin regulation of liver receptor homolog-1 mediates pluripotency gene expression. STEM CELLS *28*, 1794-1804.
Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.
Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.-H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., *et al.* (2010). Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. Cell Stem Cell *7*, 618-630.
West, J.A., Park, I.-H., Daley, G.Q., and Geijsen, N. (2006). In vitro generation of germ cells from murine embryonic stem cells. Nat Protocols *1*, 2026-2036.
West, J.A., Viswanathan, S.R., Yabuuchi, A., Cunniff, K., Takeuchi, A., Park, I.-H., Sero, J.E., Zhu, H., Perez-Atayde, A., Frazier, A.L., *et al.* (2009). A role for Lin28 in primordial germ-cell development and germ-cell malignancy. Nature *460*, 909-913.
Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810-813.
Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol *13*, 838-845.
Wu, M.-Y., Tsai, T.-F., and Beaudet, A.L. (2006a). Deficiency of Rbbp1/Arid4a and Rbbp1l1/Arid4b alters epigenetic modifications and suppresses an imprinting defect in the PWS/AS domain. Genes & Development *20*, 2859-2870.
Wu, Q., Chen, X., Zhang, J., Loh, Y.H., Low, T.Y., Zhang, W., Sze, S.K., Lim, B., and Ng, H.H. (2006b). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J Biol Chem *281*, 24090-24094.
Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N., and Tada, T. (2005). Nanog expression in mouse germ cell development. Gene Expression Patterns *5*, 639-646.
Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat Genet *40*, 1016-1022.
Yamanaka, S. (2007). Strategies and New Developments in the Generation of Patient-Specific Pluripotent Stem Cells. Cell Stem Cell *1*, 39-49.
Yamanaka, S., and Takahashi, K. (2006). [Induction of pluripotent stem cells from mouse fibroblast cultures]. Tanpakushitsu Kakusan Koso *51*, 2346-2351.
Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development *122*, 881-894.
Yi, F., Pereira, L., Hoffman, J.A., Shy, B.R., Yuen, C.M., Liu, D.R., and Merrill, B.J. (2011). Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. Nat Cell Biol *13*, 762-770.
Yi, F., Pereira, L., and Merrill, B.J. (2008). Tcf3 Functions as a Steady-State Limiter of Transcriptional Programs of Mouse Embryonic Stem Cell Self-Renewal. STEM CELLS *26*, 1951-1960.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R.*, et al.* (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science *318*, 1917-1920.
Zeng, W., Baumann, C., Schmidtmann, A., Honaramooz, A., Tang, L., Bondareva, A., Dores, C., Fan, T., Xi, S., Geiman, T.*, et al.* (2011). Lymphoid-Specific Helicase (HELLS) Is Essential for Meiotic Progression in Mouse Spermatocytes. Biology of Reproduction *84*, 1235-1241.
Zhang, J., Tam, W.-L., Tong, G.Q., Wu, Q., Chan, H.-Y., Soh, B.-S., Lou, Y., Yang, J., Ma, Y., Chai, L.*, et al.* (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. Nat Cell Biol *8*, 1114-1123.
Zhou, H., Wu, S., Joo, J.Y., Zhu, S., Han, D.W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y.*, et al.* (2009). Generation of Induced Pluripotent Stem Cells Using Recombinant Proteins. Cell Stem Cell *4*, 381-384.